Michael S. A. Graziano

CONȘTIINȚA

DINTR-O NOUĂ PERSPECTIVĂ

O teorie ştiinţifică
a experienţei subiective

Michael S.A. Graziano s-a născut în 1967 în Bridgeport, Connecticut. După studii în psihologie și științe cognitive la Universitatea Princeton și la MIT, și-a susținut doctoratul în neuroștiințe la Princeton în 1996. Tot aici duce o amplă și importantă muncă de cercetare care își propune să explice cum creierul monitorizează spațiul din jurul corpului, cum controlează mișcarea în acel spațiu și cum construiește conștiința. Descoperirile sale s-au concretizat în numeroase conferințe susținute în marile universități ale lumii și în zeci de articole științifice și de popularizare.

Este profesor de psihologie și neuroștiințe la Universitatea Princeton din 2001.

Printre cărțile sale se numără: *The Spaces Between Us*: *A Story of Neuroscience*, *Evolution*, *and Human Nature* (2018); *Consciousness and the Social Brain* (2013); *God*, *Soul*, *Mind*, *Brain*: *A Neuroscientist*'s *Reflections on the Spirit World* (2010); *The Intelligent Movement Machine*: *An Ethological Perspective on the Primate Motor Sistem* (2008). Graziano a publicat, de asemenea, romane, cărți despre muzică și cărți pentru copii, acestea din urmă sub pseudonimul B. B. Wurge.

Traducere din engleză de Vlad Vedeanu


Redactor: Anca Lăcătuș
Coperta: Ioana Nedelcu
Tehnoredactor: Manuela Măxineanu
Corector: Cristian Negoiță
DTP: Corina Roncea, Dan Dulgheru
Tipărit la Paper Print – Brăila
Michael S.A. Graziano
Rethinking Consciousness: *A Scientific Theory of Subiective Experience*
First published by W.W. Norton & Company, Inc.

Descrierea CIP a Bibliotecii Naționale a României
Graziano, Michael S.A.
Conștiința dintr-o nouă perspectivă: o teorie științifică a experienței subiective / Michael S.A. Graziano;
trad. Din engleză de Vlad Vedeanu.
— București: Humanitas, 2021 ISBN 978 – 973 – 50 – 6920 – 9
EDITURA HUMANITAS
Piața Presei Libere 1, 013701 București, România tel. 021.408.83.50, fax 021.408.83.51 www.humanitas.ro
Comenzi online: www.libhumanitas.ro
Comenzi prin email: vânzărilibhumanitas.ro
Comenzi telefonice: 0723.684.194

Pentru Ben și Ele

Cuprins

1. Elefantul din cameră
2. Crabi și caracatițe
3. Inteligența centrală a broaștei 2
4. Cortexul cerebral și conștiința
5. Conștiința socială
6. Yoda și Darth: Cum putem descoperi conștiința în creier? 8
7. Problema dificilă și alte perspective asupra conștiinței 111
8. Mașinării conștiente 142
9. Transferul minților 164
Anexă: Cum se construiește conștiința vizuală
Note

1
Elefantul din cameră

Când fiul meu avea trei ani, i-am făcut elefantul de pluș să vorbească. La vârsta lui nu avea cum să-și dea seama ce ventriloc nepriceput sunt, așa că trucul a mers foarte bine cu el. I-a plăcut mult. În următorii ani, pe măsură ce mi-am îmbunătățit tehnica, a început să mă impresioneze ciudata putere a acestei iluzii. Nu e doar o voce care aparent vine de la o păpușă ca și cum ar proveni de la un difuzor ascuns. Chiar și în mâinile unui artist mediocru ca mine se întâmplă ceva deosebit. Păpușa prinde viață, are propria personalitate și pare să aibă conștiință.

Creierul uman are desigur mecanisme care ne obligă să-i atribuim păpușii conștiință. Dar nu am dezvoltat aceste mecanisme pe parcursul evoluției ca să ne bucurăm de ventriloci. Oamenii sunt animale sociale, și ne facem adesea acest truc unii altora. Când vorbesc cu cineva, am impresia automată că acea persoană are gânduri, emoții și conștiință. Bineînțeles, nu percep direct mintea celuilalt. În schimb, creierul meu construiește un model util al unei minți și îl proiectează asupra acelei persoane, tratând-o așa cum fiul meu trata păpușa.

Nu aplicăm acest proces doar oamenilor. Atribuim conștiință câinilor sau pisicilor noastre și unii oameni ar putea să jure că plantele lor sunt conștiente. Strămoșii noștri erau siguri că râurile și copacii aveau conștiință; copiii atribuie conștiință jucăriilor lor preferate; și, la naiba, alaltăieri m-am enervat pe computerul meu. Așa că nu vorbesc despre procesul intelectual de a descoperi dacă ceva are minte sau despre operația de a deduce în moduri ingenioase care ar putea fi conținuturile acelei minți – deși facem și asta. Vorbesc despre intuiția automată, viscerală, adesea greșită, dar uneori tare convingătoare, că un obiect are conștiință.

Pe măsură ce mă gândeam la activitatea mea de

ventriloc am început să mă întreb dacă propria mea conștiință și aceste exemple de atribuire a conștiinței ar putea avea aceeași sursă. Poate că există o explicație unificatoare: construim automat modele ale minții și le proiectăm asupra propriei persoane și asupra altor oameni. Intuițiile noastre legate de o misterioasă prezență conștientă, convingerea noastră că e prezentă în mine, sau în tine, sau într-un animal de companie, sau într-un obiect s-ar putea să depindă de aceste modele simplificate, dar utile – seturi de informații pe care creierul le construiește pentru a-nțelege lumea.

Astfel de cugetări profunde nu pot să apară decât când vorbești cu un elefant de pluș. Și constituie motivul pentru care mi-am concentrat eforturile științifice asupra studiului conștiinței.

Timp de douăzeci de ani am studiat aspecte considerate tradiționale în neuroștiințe – modul în care creierul monitorizează spațiul din imediata apropiere a corpului și modul în care controlează mișcările complexe în acest spațiu.[1] Experiența în bazele practice ale neuroștiinței s-a dovedit utilă pentru construirea unei teorii a conștiinței. În 2010, colegii mei și cu mine am început să conturăm ceea ce am numit „teoria schemei atenției” (*attention schema theory*), bazându-ne pe neuroștiințe, psihologie și evoluționism și adăugând descoperiri tehnologice.[2] Această teorie face parte dintr-o mai amplă schimbare a

---

[1] Cercetările mele asupra spațiului personal și mișcărilor complexe sunt rezumate în două cărți. M.S.A. Graziano, The Intelligent Movement Machine (Oxford, UK: Oxford University Press, 2008); M.S.A. Graziano, The Spaces Between Us: A Story of Neuroscience, Evolution, and Human Nature (New York: Oxford University Press, 2018).

[2] Următoarele referințe oferă o perspectivă de ansamblu asupra teoriei. Alte surse, mai tehnice sau centrate pe date, nu sunt enumerate aici. M.S.A. Graziano și S. Kastner, „Human Consciousness and Its Relationship to Social Neuroscience: A Novel Hypothesis”, Cognitive Neuroscience 2 (2011): 98 – 113; M.S.A. Graziano, Consciousness and the Social Brain (Oxford, UK: Oxford University Press, 2013); T.W. Webb și M.S.A. Graziano, „The Attention Schema Theory: A Mechanistic Account of Subiective Awareness”, Frontiers în Psychology 6 (2015): articolul 500.

perspectivei în cadrul comunității științifice.[3] Noua abordare nu rezolvă așa-numita „problemă dificilă” (*hard problem*) a conștiinței – modul în care un creier fizic poate genera o esență non-fizică.[4] În schimb, explică de ce oamenii pot considera în mod greșit că există o problemă dificilă, de ce această intuiție greșită este profund înrădăcinată în noi, astfel că e puțin probabil să o putem schimba, și de ce prezența ei este avantajoasă și poate chiar necesară pentru funcționarea creierului.

La început am privit teoria din perspectiva interacțiunii sociale. Însă, la origine, această teorie depinde de o proprietate mai degrabă generală a creierului: cunoașterea bazată pe modele.[5] Creierul construiește modele interne – pachete bogate de informații, aflate permanent în schimbare, structurate continuu și automat, ca niște bule de sensuri situate sub nivelul superior al gândirii sau limbajului. Aceste modele interne înfățișează elemente importante, a căror monitorizare este utilă: uneori obiecte din exterior, alteori aspecte ale sinelui. Reprezentările sunt simplificate și distorsionate, ca niște desene impresioniste sau cubiste ale realității, și ne raportăm la conținuturile lor ca și cum ar reflecta realitatea concretă. Nu putem să nu facem asta – așa suntem construiți. Modul în care înțelegem intuitiv lumea din jur și propria persoană, întotdeauna distorsionat și simplificat, depinde de aceste modele interne.

---

[3] Este imposibil să prezentăm aici tot ce s-a scris recent despre abordările mecaniciste, nondualiste ale conștiinței. Dau câteva exemple și le cer iertare numeroșilor autori excelenți pe care nu i-am inclus. S.J. Blackmore, „Consciousness în Meme Machines”, Journal of Consciousness Studies 10 (2003): 19 – 30; P.S. Churchland, Touching a Nerve: Our Brains, Our Selves (New York: W.W. Norton, 2013); F. Crick, The Astonishing Hypothesis: The Scientific Search for the Soul (New York: Scribner, 1995);

[4] Chalmers, „Facing Up to the Problem of Consciousness”, Journal of Consciousness Studies 2 (1995): 200 – 219.

[5] O abordare mai veche și perspicace a conștiinței care pune accentul pe modelele interne este: O. Holland și R. Goodman, „Robots with Internal Models: A Route to Machine Consciousness?”, Journal of Consciousness Studies 10 (2003): 77 – 109.

În cadrul acestei teorii, intuițiile noastre metafizice referitoare la noi înșine, la conștiință ca esență interioară non-fizică – uneori numită „fantoma din mașinărie”[6] – sunt derivate dintr-un anumit model intern. Este ceea ce eu numesc schemă a atenției, din motive care se vor limpezi pe parcursul acestei cărți. E o imagine simplificată a modului în care creierul captează informația și o procesează în profunzime. Această imagine constituie un mod eficient prin care creierul înțelege și monitorizează propriile capacități interne. Același tip de model intern poate fi folosit, în mai mică măsură, pentru a monitoriza și a face predicții în legătură cu alți oameni.

Ar putea părea că abordarea centrată pe modele anulează sau devalorizează conștiința – ceea ce nu e deloc adevărat. Modelul intern care ne spune că suntem conștienți este profund, bogat, continuu și, probabil, necesar. Aproape nimic din ceea ce facem – indiferent că e vorba de percepție, gândire, acțiune, interacțiune socială – nu ar funcționa fără această parte a sistemului.

În cartea de față voi folosi interșanjabil termenii *conștiință*, *conștiință subiectivă* și *experiență subiectivă*, deși știu că aceste cuvinte nu sunt utilizate întotdeauna cu același sens de oamenii de știință. În special cuvântul *conștiință* este cunoscut pentru numeroasele sale conotații alunecoase. Înainte să spun ce înțeleg prin acesta aș vrea să clarific la ce *nu* mă refer. Uneori oamenii consideră conștiința capacitatea de a ști cine ești și de a-ți înțelege traiectoria propriei vieți. Alții o consideră mai degrabă capacitatea de a procesa lumea din jur și de a lua decizii în consecință. Eu mă refer la cu totul altceva.

Cel mai comun sens al experienței interne este probabil „fluxul conștiinței”, conținutul caleidoscopic al minții. Este vorba de tumultul de gânduri pe care James Joyce l-a surprins în faimosul lui roman din 1922, *Ulise*.[7] Joyce a

---

[6] G. Ryle, The Concept of Mind (Chicago: University of Chicago Press, 1949).

[7] J. Joyce, Ulysses (Paris: Sylvia Beach, 1922). [V. Ed. Rom.: Ulise, trad.

înregistrat cu meticulozitate senzațiile vizuale, și auditive, și tactile, aflate în permanentă schimbare, pe care lumea le transmite – gusturi și mirosuri, reveniri ale amintirilor recente și de demult, dialogul intern în continuă desfășurare, emoțiile și fanteziile conflictuale, dintre care unele atât de scandaloase, încât cartea a fost inițial interzisă. (Procesul din 1933 „Statele Unite ale Americii vs. Cartea intitulată Ulise” ne-a oferit definiția juridică modernă a obscenității.) Dar, din nou, nu asta vreau să spun prin conștiință. Acest flux de material nu este foarte bine definit, iar volumul său enorm îl face imposibil de studiat științific. Imaginează-ți, în schimb, că pui 10.000 de lucruri într-o găleată. Poți să le cataloghezi într-o listă complicată de obiecte, cum a făcut Joyce. Dar ai putea și să-ți pui o întrebare fundamentală: care-i treaba cu găleata? Lasă conținutul deocamdată. Din ce e făcută găleata, de unde vine? Cum ajunge un om să fie conștient de orice? Conștiința nu poate fi doar informația din noi, dat fiind că nu suntem conștienți decât de o mică parte din baza enormă de informații stocată la un moment dat în creierul nostru. Ceva trebuie să se petreacă pentru ca un anumit volum de informație să ajungă în conștiință. Cum se întâmplă asta? Această întrebare mai specifică i-a preocupat din ce în ce mai mult pe filozofi şi pe oamenii de ştiinţă[8]. Termenul de conştiinţă a ajuns să însemne actul de a fi conștient de ceva, și nu materialul de care ești conștient.

Presupun că schimbarea treptată, în filosofie, a accentului de la numeroasele elemente din fluxul conștiinței la actul de *a fi conștient* este legată în mare măsură de progresul computerelor din ultima jumătate de secol. Pe măsură ce tehnologia informației s-a îmbunătățit, conținuturile informaționale ale minții au devenit mai puțin misterioase și, în același timp, actul de a fi conștient de acestea, de a avea orice fel de experiență, a devenit mai

De Mircea Ivănescu, Editura Univers, București, 2012.]

[8] S. Koenemann și R. Jenner, *Crustacea and Arthropod Relationships* (Boca Raton: CRC Press, 2005).

îndepărtat și aparent de nerezolvat. Să luăm câteva exemple.

Poți să conectezi o cameră digitală la un computer și să programezi sistemul să proceseze informația vizuală descărcată. Computerul poate extrage culoarea, forma și dimensiunea și poate identifica obiecte. Creierul uman face ceva similar. Diferența este că oamenii au și o experiență subiectivă a ceea ce văd. Nu înregistrăm pur și simplu informația că un obiect e roșu; avem *experiența* roșului. Să vezi ceva *se simte* într-un fel anume. Un computer modern poate procesa o imagine vizuală, dar specialiștii încă nu au descoperit cum să facă computerul să fie conștient de acea informație.

Acum gândește-te la ceva puțin mai personal decât percepția vizuală: amintirile autobiografice care-ți definesc traiectoria în viață. Nimic nu exemplifică mai bine fluxul joycean al conștiinței decât amintirile care apar constant în minte. Și totuși știm cum să construim o mașină care are memorie, care stochează și reactualizează informații. Orice computer are această capacitate, și oamenii de știință cunosc principiile generale, dacă nu chiar și detaliile, modului în care amintirea este stocată în creier. Memoria nu e un mister fundamental. Și nici nu determină conștiința. Din nou, conținutul *din* conștiință – în acest exemplu, o amintire – nu e același lucru cu actul de *a fi conștient de* o amintire.

O să mai dau un exemplu: luarea deciziilor. Dacă e ceva ce definește misterul conștiinței umane, cu siguranță trebuie să fie capacitatea noastră de a lua decizii. Asimilăm informații, le procesăm, le judecăm și alegem ce să facem în continuare. Dar, iarăși, aș spune că conștiința nu este o parte intrinsecă a luării de decizii. Toate computerele iau decizii. Într-un sens, asta e definiția computerului. Asimilează informații, le manipulează și le utilizează pentru a selecta una dintre numeroasele căi de acțiune. Majoritatea deciziilor pe care le ia creierul uman, poate zeci de mii în fiecare zi, au loc automat, fără nicio experiență subiectivă. În unele cazuri rare avem conștiința subiectivă a

luării unei decizii. Uneori o numim intenție, alegere sau liber-arbitru. Dar capacitatea în sine de a lua o decizie nu necesită conștiință.

Prin aceste exemple și prin multe altele, progresul tehnologiei informatice a dezvăluit diferența dintre conținutul conștiinței, din ce în ce mai bine înțeles din perspectivă tehnică, și actul de a fi conștient de acesta. Pe mine mă interesează această a doua parte, esențială, a problemei: cum ajungem să avem o experiență subiectivă de orice fel?

Uneori oamenii consideră limitativă această abordare. Am fost întrebat adesea: cum rămâne cu memoria? Cum rămâne cu alegerea conștientă? Cum rămâne cu înțelegerea de sine? Cum rămâne cu intențiile și convingerile? Nu sunt ele elementele fundamentale ale conștiinței? Sunt de acord: toate acestea sunt importante și reprezintă elemente esențiale ale conștiinței umane. Cu toate astea, nu sunt mistere fundamentale. Sunt chestiuni legate de procesarea informației și putem să ne imaginăm, măcar în principiu, modul în care le putem construi sau modifica. Misterul fundamental este tocmai găleata. Ce e conștiința – din ce e făcută? Cum poate ceva să ajungă în conștiință, care e câștigul dacă ajunge acolo și de ce atât de puține elemente din creier ajung acolo?

În mod tradițional, oamenii de știință au presupus că ceva atât de amorf și alunecos cum e conștiința trebuie să fie imposibil de înțeles științific. Dar, date fiind descoperirile recente, acum sunt destul de sigur că poate fi înțeleasă și construită în aceeași măsură în care pot fi înțelese și construite procesarea vizuală, memoria, luarea de decizii sau oricare alt element pe care îl conține.

*

Am scris de multe ori până acum despre conștiință. Însă această carte este destinată întru totul cititorului nespecialist. Încerc aici să expun cât se poate de simplu și clar o teorie științifică promițătoare a conștiinței – o teorie care poate fi aplicată în aceeași măsură creierului biologic

și dispozitivelor înzestrate cu inteligență artificială.

Următoarele câteva capitole pornesc de la evoluție. Începând cu peste jumătate de miliard de ani în urmă, de la apariția neuronilor – celulele care alcătuiesc creierul –, voi descrie complexitatea din ce în ce mai mare a sistemului nervos. Pe parcurs, voi introduce elementele teoriei schemei atenției și până la Capitolul 6 va fi pusă temelia acesteia.

Voi discuta apoi modul în care teoria schemei atenției se leagă de celelalte teorii. Teoria schemei atenției este una dintre cele câteva teorii importante ale conștiinței care câștigă teren în literatura științifică. Impresia mea, pe care încerc să o transmit în această carte, este că aceste teorii nu ar trebui considerate rivale și că nu e cazul să așteptăm să vedem care din ele va ieși câștigătoare. Oricât ar fi de diferite – și nu sunt de acord cu multe dintre ideile propuse –, aceste teorii numeroase pot avea legături ciudate și ascunse unele cu celelalte. Fiecare contribuie cu descoperiri importante. Cred că începem să vedem licărirea unui consens – sau poate mai degrabă o rețea de idei aflate în acord unele cu altele.

În capitolele de final voi aprofunda implicațiile tehnologice. Suntem destul de aproape să-nțelegem conștiința suficient de bine încât să o putem construi, și, când vom face asta, probabil că noua tehnologie ne va schimba complet civilizația. Conștiința artificială este doar primul pas. Dacă conștiința poate fi construită, atunci mintea este, în principiu, transferabilă de la un dispozitiv la altul. Deși într-un viitor mult mai îndepărtat, teoretic este posibilă citirea informațiilor relevante din creierul uman și transferarea minții unei persoane într-o platformă artificială.[9] Tehnologia i-ar putea permite minții să trăiască

[9] R.A. Koene, „Scope and Resolution în Neural Prosthetics and Special Concems for the Emulation of a Whole Brain”, Journal of Geoethical Nanotechnology 1 (2006): 21 – 29; R. Kurzweil, The Singularity Is Ne-ar: When Humans Transcend Biology (New York: Penguin Books, 2006); H. Markram, E. Müller, S. Ramaswamy, M.W. Reimann, M. Abdellah, C.A. Sánchez, A. Ailamaki et al., „Reconstruction and Simulation of Neocortical Microcircuitry”, Cell 163 (2015): 456 – 92; A. Sandberg și N.

oricât și să exploreze medii ostile corpului biologic, cum ar fi spațiul interplanetar. Legile fizicii nu ne stau în cale – trebuie doar inventate dispozitivele potrivite.

Dacă e posibilă înțelegerea conștiinței dintr-o perspectivă științifică și tehnică, atunci acest subiect nu mai este doar un joc filosofic pentru oamenii de știință. Devine o chestiune practică urgentă. Cartea de față va expune modurile în care conștiința va putea fi întrebuințată în viitor, unele bune, altele de-a dreptul odioase. Dar, bune sau rele, sunt destul de sigur că ne apropiem rapid de înțelegerea științifică a conștiinței și de momentul în care o vom putea construi artificial.

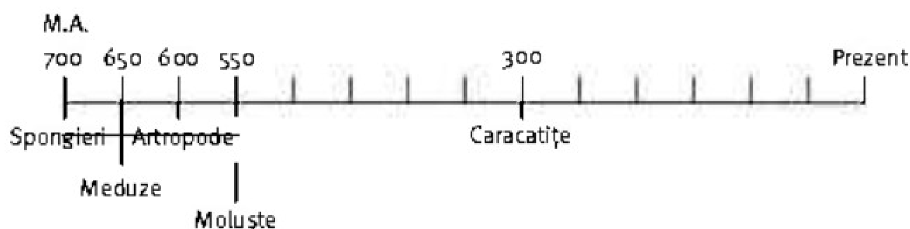
2
Crabi și caracatițe

Celulele bacteriene au apărut pentru prima oară pe Pământ în urmă cu aproximativ patru miliarde de ani. În cea mai mare parte a istoriei planetei, viața a rămas la stadiul unicelular și nu a existat nimic asemănător unui sistem nervos până în urmă cu vreo 600 sau 700 de milioane de ani (m.a.). În teoria schemei atenției, conștiința depinde de un sistem nervos care să proceseze informația într-un anumit mod. Cheia teoriei, și presupun că a oricărei inteligențe avansate, este atenția – capacitatea creierului ca într-un anumit moment să-și concentreze resursele limitate asupra unui anumit fragment din lume pentru a-l procesa în profunzime. În acest capitol și în cele următoare voi examina modul în care se poate ca atenția să fi evoluat de la primele animale la oameni și modul în care se poate ca proprietatea pe care o numim conștiință să fi apărut odată cu această tranziție.[0]

Bostrom, „Whole Brain Emulation: A Roadmap”, Tehnical Report 2008 – 3, Future of Humanity Institute, Oxford University, 2008.

[0] Alți autori au prezentat cu măiestrie posibila evoluție a conștiinței, precum și legătura dintre conștiință și atenție, deși dintr-o altă perspectivă decât a mea. De exemplu: C. Montemayor și H.H. Haladjian, Consciousness, Attention, and Conscious Attention

Voi începe povestea cu bureții-de-mare, spongierii fiind potriviți pentru a exemplifica evoluția sistemului nervos. Aceștia sunt cele mai primitive dintre animalele pluricelulare, fără caracteristici morfologice distincte, fără membre, fără mușchi și fără să aibă nevoie de nervi. Stau pe fundul oceanului și filtrează ca o sită substanțele nutritive. Și totuși o parte dintre genele noastre sunt aceleași cu ale spongierilor, printre care cel puțin 25 care la oameni contribuie la structurarea sistemului nervos.[0] La spongieri, aceste gene ar putea fi implicate în aspecte mai simple ale modului în care celulele comunică între ele. Bureții-de-mare par să reprezinte pragul evoluționist al sistemului nervos. Se consideră că au cel puțin un strămoș comun cu al oamenilor, care a viețuit cândva în urmă cu 700 sau 600 de milioane de ani (vezi Figura 2.1).[0]



2.1 Nevertebratele despre care discutăm în acest capitol și perioada aproximativă în care au apărut.

Spre deosebire de spongieri, un alt tip străvechi de animal, meduza, are sistem nervos. Meduzele nu se fosilizează prea bine, dar, analizând relația genetică pe care o au cu alte animale, biologii estimează că se poate să

(Cambridge MA: MIT Press, 2015); R. Ornstein, Evolution of Consciousness: The Origins of the Way We Think (New York: Simon and Schuster, 1991).

[0] O. Sakarya, K.A. Armstrong, M. Adamska, M. Adamski, I.F. Wang, B. Tidor, B.M. Degnan, T.H. Oakley, și K.S. Kosik, „A Post-Synaptic Scaffold at the Origin of the Animal Kingdom”, PLOS One 2 (2007): e506.

[0] Z. Yin, M. Zhu, E.H. Davidson, D.J. Bottjer, F. Zhao și P. Tafforeau, „Sponge Grade Body Fossil with Celular Resolution Dating 60 Myr Before the Cambrian”, Proceedings of the National Academy of Sciences U.S.A. 112 (2015): E1453 – 1460.

se fi diferențiat de restul regnului animal începând cu 650 de milioane de ani în urmă.[0] Aceste numere se pot schimba în funcție de noi informații, dar ca estimare plauzibilă se pare că neuronii, celulele de bază ale sistemului nervos, au apărut pentru prima oară în regnul animal undeva între spongieri și meduze, cu puțin mai mult de o jumătate de miliard de ani în urmă.

Neuronul este, în esență, o celulă care transmite un semnal. Un val de energie electrochimică străpunge celula dintr-o parte a membranei în cealaltă cu o viteză de aproximativ 61 m/s și influențează un alt neuron, un mușchi sau o glandă. Se poate ca primele sisteme nervoase să fi fost rețele simple de neuroni care se întindeau de-a lungul corpului și legau mușchii. Hidrele funcționează pe acest principiu de rețea nervoasă.[0] Acestea sunt mici creaturi marine – animale transparente care seamănă cu florile, cu corpul în formă de sac și multe tentacule – și aparțin aceleiași categorii străvechi din care fac parte meduzele. Dacă atingi hidra într-un punct, rețeaua nervoasă transmite semnalul în mod nediferențiat, și aceasta va tremura din tot corpul.

O rețea nervoasă nu procesează informația – nu într-un sens important. Nu face decât să transmită semnale în corp. Face legătura între un stimul senzorial (atingerea hidrei) și un răspuns muscular (tremorul). Însă, după apariția rețelei nervoase, sistemele nervoase au evoluat rapid la un al doilea nivel de complexitate: capacitatea de-a amplifica unele semnale în defavoarea altora. Acest truc simplu, dar important, de amplificare a semnalului este

[0] D.H. Erwin, M. Laflamme, S.M. Tweedt, E.A. Sperling, D. Pisani și KJ. Peterson, „The Cambrian Conundrum: Early Divergence and Later Ecological Succes în the Early History of Animals”, Science 334 (2011): 1091 – 1097; A.C. Marques și A.G. Collins, „Clădistic analysis of Medusozoa and cnidarian evolution”, Invertebrate Biology 123 (2004): 23 – 42.

[0] H.R. Bode, S. Heimfeld, O. Koizumi, C.L. Littlefield și M.S. Yaross, „Maintenance and Regeneration of the Nerve Net în Hydra”, American Zoology 28 (1988): 1053 – 1063.

unul dintre modurile fundamentale în care neuronii manipulează informația. Este o componentă a aproape tuturor tipurilor de calcul pe care știm că le face creierul.

Ochiul de crab este unul din exemplele cele mai studiate.[0] Crabul are ochi compuși, cu o mulțime de detectori, fiecare având un neuron. Contactul cu lumina activează neuronul dinăuntrul detectorului. Până aici totul e bine. Dar, lucrurile fiind puțin mai complexe, fiecare neuron este conectat la cei din imediata apropiere și, dată fiind această conexiune, neuronii concurează între ei. Când se activează un neuron dintr-un detector, acesta are tendința de a suprima activitatea neuronilor din detectorii din vecinătate, ca un om care strigă cât îl țin puterile într-o mulțime ca să îi acopere pe ceilalți.

Rezultatul este că, dacă cea mai strălucitoare parte a unui punct difuz de lumină ajunge la un detector din ochiul crabului, neuronul din acest detector se activează, își învinge competitorii și întrerupe activitatea celorlalți. Acest tipar de activitate de la nivelul detectorilor vizuali nu numai că semnalează un punct luminos, ci face să apară și un cerc de întuneric în jurul acestuia. Semnalul este astfel amplificat. Ochiul crabului face din realitatea estompată, gri, o imagine cu contrast puternic, cu alb foarte deschis și negru foarte închis. Această amplificare a semnalului este o consecință directă a faptului că neuronii inhibă activitatea neuronilor din apropiere, proces numit *inhibiție laterală*.[0]

Mecanismul operant la nivelul ochiului de crab este probabil exemplul cel mai simplu și fundamental – prototipul – atenției. Semnalele concurează unele cu altele, cele care câștigă competiția fiind amplificate în defavoarea celorlalte și ajungând să influențeze mișcările animalului. *Aceasta* e esența computațională a atenției. Atenția oamenilor este doar o versiune elaborată a acesteia, constituită din aceleași elemente. Mecanismul inhibiției

[0] R.B. Barlow Jr. Și A.J. Fraioli, „Inhibition în the Limulus Lateral Eye în Situ”, Journal of General Physiology 71 (1978): 699 – 720.

[0] K. Hadeler, „On the Theory of Lateral Inhibition”, Kybernetik 14 (1974): 161 – 165.

laterale prezent la crab poate fi găsit în orice stadiu de procesare al sistemului nervos al oamenilor, de la ochi la cele mai înalte niveluri de gândire care se desfășoară în cortexul cerebral. Originea atenției se află departe pe scara evoluției, cu mai bine de jumătate de miliard de ani în urmă, și a apărut printr-o inovație surprinzător de simplă.

Crabii fac parte dintr-un grup mare de animale, artropodele, care include păianjenii și insectele și alte creaturi cu exoschelet dur și articulat și care au evoluat din alte animale cu aproximativ 600 de milioane de ani în urmă.[0] Cel mai celebru artropod acum dispărut, cel care astăzi are cel mai mare grup de fani, este trilobitul, o creatură cu multe picioare care seamănă cu un crab-potcoavă în miniatură și care se târa pe fundul mării în perioada cambrian, în urmă cu 549 de milioane de ani. Când trilobiții au murit și s-au scufundat în depunerile de pe fundul oceanului, ochii lor fațetați s-au fosilizat uneori, păstrând detalii uluitoare.[0] Dacă te uiți la o fosilă de trilobit și îi examinezi ochii proeminenți printr-o lupă, adesea poți vedea încă mozaicul bine ordonat al detectorilor. Judecând pe baza acestor detalii fosilizate, ochii trilobiților trebuie să fi semănat îndeaproape cu ochii crabilor din ziua de azi în privința organizării, și este probabil să fi folosit același truc al competiției între detectori învecinați pentru a vedea mai bine ce se afla pe fundul mării.

Să ne imaginăm un animal construit bucată cu bucată cu atenție „locală”. Fiecare parte a corpului ar funcționa ca un aparat separat, filtrând propriile informații și recepționând semnalele cele mai accentuate. Un ochi poate să spună: „Punctul ăsta e foarte luminos. Să le ignorăm pe celelalte”. În același timp, în mod independent, un picior

[0] S. Koenemann și R. Jenner, Crustacea and Arthropod Relationships (Boca Raton: CRC Press, 2005).
[0] B. Schoenemann, H. Pămaste și E.N.K. Clarkson, „Structure and Function of a Compound Eye, More Than Half a Billion Years Old”, Proceedings of the National Academy of Sciences U.S.A. 114 (2017): 13489 – 13494.

spune: „Tocmai am primit o lovitură aici. Ignoră alte atingeri mai ușoare!” Un animal care nu ar funcționa decât așa ar fi ca un grup de agenți separați care se întâmplă să fie lipiți laolaltă, fiecare strigându-și propriul semnal, declanșând propriile acțiuni. Comportamentul animalului ar fi, în cel mai bun caz, haotic.

Pentru a avea o reacție coerentă la mediu, un animal are nevoie de atenție centralizată. E oare posibil ca numeroasele surse de imput – ochii, corpul, picioarele, urechile, senzorii chimici – să adune informațiile într-un singur loc pentru o evaluare globală și o ierarhizare a semnalelor? Această convergență i-ar permite animalului să aleagă obiectul cel mai pregnant din mediu, cel care pare cel mai important într-un anumit moment, și apoi să genereze un răspuns unic, semnificativ.

Nimeni nu știe când a apărut prima dată acest tip de atenție centralizată, în parte pentru că nimeni nu știe cu siguranță care animale îl au și care nu. Vertebratele au un procesor central al atenției, pe care îl voi prezenta în capitolul următor. Dar mecanismele atenției nu au fost încă studiate îndeaproape la nevertebrate. Multe tipuri de animale, cum ar fi viermii inelați și melcii, nu au un creier central, ci au, peste tot prin corp, grupuri de neuroni sau ganglioni care realizează operații locale.[0] Probabil că aceștia nu au atenție centralizată.

Artropodele, cum ar fi crabii, insectele și păianjenii, au mai multe șanse să dețină atenție centralizată. Au un creier central sau, cel puțin, au, în zona capului, un agregat de neuroni mai mare decât cele din corp.[0] Acest ganglion mare se poate să fi evoluat în parte datorită necesităților vizuale. Dat fiind că ochii sunt în cap și că văzul este simțul cel mai complicat și cel mai sensibil la informație, capul a dezvoltat cei mai mulți neuroni. Aspecte ca mirosul, gustul, auzul și

---

[0] R. Gillette și J.W. Brown, „The Sea Slug, Pleurobranchaea califormica: A Signpost Species în the Evolution of Complex Nervous Sistems and Behavior”, Integrative and Comparative Biology 55 (2015): 1058 – 1069

[0] C.R. Smarandache-Wellmann, „Arthropod Neurons and Nervous Nystem”, Current Biology 26 (2016): R960-R965.

simțul tactil converg la rândul lor în acest ganglion central. Insectele sunt mai deștepte decât credem îndeobște. Când dai să plesnești o muscă și reușește să scape – cum se întâmplă mai mereu –, la mijloc nu e un simplu reflex. Musca are probabil ceva ce putem numi atenție centrală sau capacitatea de a-și concentra rapid resursele de procesare asupra oricărui aspect al lumii care pare mai important în acel moment, pentru a genera un răspuns coordonat.[0]

Caracatițele sunt superstaruri între nevertebrate datorită inteligenței lor uimitoare. Sunt considerate moluște, la fel ca scoicile sau melcii. Moluștele au apărut probabil cu 550 de milioane de ani în urmă și au rămas relativ simple, cel puțin în privința organizării sistemului nervos, timp de sute de milioane de ani.[0] Dintre încrengăturile acestui regn, cefalopodele au dezvoltat un creier complex și un comportament sofisticat și au ajuns să semene întrucâtva cu forma pe care o au azi în urmă cu aproximativ 300 de milioane de ani.[0]

Caracatițele, calamarii și sepiile sunt extratereștri în raport cu noi.[0] Nici un alt animal inteligent nu este mai

---

[0] S. Koenig, R. Wolf și M. Heisenberg, „Visual Attention în Flies-Dopamine în the Mushroom Bodies Mediates the After-Effect of Cueing”, PLOS One 11 (2016): eol 61412; B. Van Swinderen, „Attention în Drosophila”, International Review of Neurobiology 99 (2011): 51 – 85.

[0] S. Koenig, R. Wolf și M. Heisenberg, „Visual Attention în Flies-Dopamine în the Mushroom Bodies Mediates the After-Effect of Cueing”, PLOS One 11 (2016): eol 61412; B. Van Swinderen, „Attention în Drosophila”, International Review of Neurobiology 99 (2011): 51 – 85.

[0] J. Kluessendorf și P. Doyle, „Pohâsepia mazonensis, An Early Octopus’ From The Carboniferous Of Illinois, USA”, Palaeontology 43 (2000): 919 – 926; A.R. Tanner, D. Fuchs, le. Winkelmann, M.T. Gilbert, M.S. Pankey, A.M. Ribeiro, K.M. Kocot, K.M. Halanych, T.H. Oakley. R. R. Da Fonseca, D. Pisani și J. Vinther, „Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution”, Proceedings of Royal Society, B, Biological Sciences 284 (2017):

[0] P. Godfrey-Smith, Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness (New York: Farrar, Straus and Giroux, 2016); S. Montgomery, The Soulofan Octopus (New York: Atria Books,

departe de noi pe Pământ. Ne arată că inteligența determinată de un creier mare nu este un eveniment unic, dat fiind că a evoluat independent de cel puțin două ori – mai întâi la vertebrate și apoi din nou la nevertebrate.

Caracatițele sunt prădători excelenți, care se bazează pe simțul vizual. Un bun prădător trebuie să fie mai inteligent și mai bine coordonat decât prada, iar utilizarea văzului pentru localizarea și recunoașterea prăzii este un proces deosebit de intens. Nici un alt sistem senzorial nu are un volum atât de variat de informație absorbită și o nevoie atât de mare de a folosi subseturile de procesare a informației într-un mod inteligent. Așadar, atenția este esențială pentru un prădător care se bazează pe văz atunci când vânează. Poate că stilul de viață are ceva de-a face cu dezvoltarea inteligenței caracatiței.

Oricare ar fi motivul, caracatița a dezvoltat un sistem nervos extraordinar. Poate utiliza instrumente, poate rezolva probleme și prezintă o creativitate neașteptată.[0] Într-o demonstrație care a devenit clasică, caracatițele pot învăța să deschidă un borcan răsucind capacul pentru a ajunge la hrană. Caracatița are un creier central și mai are și câte un procesor independent, mai mic, la nivelul fiecărui tentacul, ceea ce o face să aibă un tip hibrid de comandă centralizată și localizată.[0] Caracatița are probabil și modele ale sinelui – pachete bogate de informații actualizate constant prin care își monitorizează corpul și comportamentul. Din punct de vedere tehnic, îi sunt necesare modele ale sinelui pentru a funcționa eficient. De exemplu, s-ar putea să aibă un tip de schemă corporală care înregistrează forma și structura propriului corp pentru

---

2015).

[0] A.-S. Darmaillacq, L. Dickel și J.A. Mather, Cephalopod Cognition (Cambridge, UK: Cambridge University Press, 2014); D.B. Edelman, B.J. Baars și A.K. Seth, „Identifying Hallmarks of Consciousness în Nonmammalian Species”, Conscious and Cognition 14 (2015): 169 – 187; f. N. Richter, B. Hochner și M.J. Kuba, „Pull or Push? Octopuses Solve a Puzzle Problem”, PLOS One 11 (2016): eO 152048.

[0] B. Hochner, „An embodied view of octopus neurobiology”, Current Biology 22 (2012): R887 – 892.

a coordona mișcările. (Poate că fiecare tentacul are propria schemă.) în acest sens, s-ar putea spune că o caracatiță știe că există. Deține informații despre sine și despre lumea din exterior, iar aceste informații duc la comportamente complexe.

Dar toate aceste trăsături cu adevărat minunate nu înseamnă că e conștientă.

Cercetătorii conștiinței folosesc uneori sintagma *conștiință obiectivă* pentru a desemna faptul că informația a fost asimilată și procesată într-un mod care influențează alegerile comportamentale.[0] Cu această definiție limitată, s-ar putea spune că un cuptor cu microunde este conștient de setarea de timp și că o mașină cu inteligență artificială este conștientă de un obstacol din apropiere. Da, caracatița este conștientă obiectiv de sine și de obiectele din jurul ei. Conține aceste informații.

Dar are conștiință *subiectivă*? Dacă ar putea să vorbească, ar spune că are o experiență subiectivă, conștientă, ca mine sau ca tine?

S-o întrebăm chiar pe ea. Să facem un experiment imaginar întrucâtva improbabil – și să-l ținem minte, pentru că ne va fi de folos pe parcursul cărții. Să presupunem că am pus mâna pe un dispozitiv științifico-fantastic incredibil – să-i spunem Vorbinator 5.000 – care transformă informația în limbaj. Are o mufă care poate fi conectată la capul caracatiței și verbalizează informațiile pe care le găsește în creier.

Ar putea face afirmații precum: „Acolo e un pește”, dacă sistemul vizual al caracatiței ar conține informații despre un pește care se află în apropiere. Ar putea spune: „Sunt o entitate cu multe membre, care se mișcă în mai multe feluri”. Ar putea spune: „Ca să scot un pește dintr-un borcan trebuie să învârt partea aia circulară”. Ar putea

[0] P.M. Merikle, D. Smilek și J.D. Eastwood, „Perception without awareness: Perspectives from cognitive psychology”, Cognition 79 (2001): 115 – 134; R. Szezepanowski și L. Pessoa, „Fear Perception: Can Obiective and Subiective Awareness Measures Be Dissociated?”, Journal of Vision 10 (2007): 1-17.

spune multe lucruri care să reflecte informațiile despre care știm că se află în sistemul nervos al caracatiței. Dar nu știm dacă ar spune: „Am o experiență subiectivă, intimă, sunt conștient de acel pește. Nu doar procesez o informație, ci am o *experiență*. Când văd un pește, am o anumită *senzație*”. Nu știm dacă creierul ei conține acest tip de informație pentru că nu știm ce îi spune modelul pe care îl are despre sine. S-ar putea să-i lipsească mecanismul prin care să creeze un model al conștiinței sau prin care să-și atribuie sieși această proprietate. S-ar putea ca pentru acest animal conștiința să fie irelevantă.

Enigma caracatiței este un exemplu instructiv al modului în care un animal poate fi complex și inteligent, dar, cu toate astea, nu putem spune până în prezent dacă are experiență subiectivă sau dacă o astfel de noțiune ar avea vreun sens pentru el.

Poate că o sursă de confuzie este aici tendința automată și puternică a oamenilor de a atribui conștiință obiectelor din jur. După cum am arătat în Capitolul 1, suntem predispuși să ne închipuim că au conștiință păpușile sau alte obiecte la care existența conștiinței e și mai neverosimilă. Oamenii cred uneori că plantele lor sunt conștiente. O caracatiță, cu comportamentele ei complexe și cu ochii ei mari și expresivi, ne testează și mai puternic, declanșând în noi o puternică percepție socială. Nu doar că știm, în plan intelectual, că asimilează informații obiective despre propria lume, însă nu ne putem abține să nu avem senzația că trebuie să aibă și o conștiință subiectivă, după cum sugerează ochii aceia care par să-i oglindească sufletul. Dar adevărul e că nu știm, și senzația pe care o avem când ne gândim la mintea ei conștientă spune mai multe despre *noi* decât despre caracatiță. Experții care studiază caracatițe riscă în acest punct să fie cel mai puțin credibili observatori, dat fiind că sunt cei mai predispuși să fie fermecați de aceste creaturi minunate. Mai târziu, în Capitolul 5, voi reveni la acest aspect universal al conștiinței umane – modul în care o folosim ca instrument din trusa noastră socială și în care o atribuim automat

agenților din jurul nostru.

Ca să fie limpede, nu spun că *nu* sunt conștiente caracatițele. Dar sistemul nervos al caracatiței nu este încă înțeles pe deplin și nu putem compara organizarea creierului ei cu a noastră sub aspectul algoritmilor și al modelelor sinelui. Pentru a face aceste comparații va trebui să examinăm animale din încrengătura noastră, cea a vertebratelor.

3
Inteligența centrală a broaștei

Mi-am petrecut o parte din copilărie la o fermă din nordul statului New York. În fiecare vară, cât era noaptea de lungă auzeam un broscoi care își orăcăia cântecul de împerechere în iazul din spatele casei. Noi îi spuneam Elvis și am botezat Priscilla vocea mai subțire care îi răspundea. Încă de atunci îmi plac broaștele și de când am devenit specialist în neuroștiințe mă interesează să aflu ce se petrece în capul lor.

Broasca are o parte a creierului numită *tectum*, care înseamnă „acoperiș” în latină; această formațiune este cea mai evidentă proeminență din partea superioară a creierului. Broaștele nu sunt singurele care au acest segment de creier. Tectumul a fost studiat cu precădere la amfibieni, dar este prezent și la pești, reptile, păsări și mamifere. Din câte știm, toate vertebratele au tectum, spre deosebire de celelalte animale. Putem face presupunerea întemeiată că, în urmă cu circa o jumătate de miliard de ani, un pește mic din clasa vertebratelor fără maxilar inferior, strămoșul vertebratelor, a dezvoltat un tectum, moștenit de toți urmașii lui.[0]

[0] E. Knudsen și f. S. Schwartz, „The Optic Tectum, a Structure Evolved for Stimulus Selection”, în J. Kaas (coord.), Evolution of Nervous Sistems (San Diego: Academic Press, 2017), 387 – 408; C. Maximino, „Evolutionary Changes în the Complexity of the Tectum of Nontetrapods: A Clădistic Approach”, PLOS One 3 (2008): e3582.

Oamenii au tectum, dar al nostru nu mai este situat în partea de sus a creierului. Este o protuberanță relativ mică – sau, mai degrabă, este alcătuit din două protuberanțe, fiecare de câte o parte a liniei mediane a creierului –, acoperită de structuri cerebrale care s-au dezvoltat pe parcursul evoluției noastre. La oameni și la alte mamifere se numește în general „colicul superior” (care în latină înseamnă „protuberanța de sus”). Aici, de dragul simplității, îi voi spune tectum.

În cea mai mare parte a evoluției vertebratelor, tectumul a reprezentat culmea dezvoltării intelectuale – cel mai complex și mai sofisticat procesor, situat în mijlocul creierului. La broască, tectumul ia informația vizuală și aranjează lumea sub forma unei hărți.[0] Fiecare punct de pe suprafața rotundă a tectumului corespunde unui punct din spațiul din jurul animalului. Partea dreaptă a tectumului broaștei conține o hartă clară, bine aranjată a câmpului vizual al ochiului stâng; la fel partea stângă pentru ochiul drept. Dacă un punct negru se mișcă haotic în jurul broaștei, ochii preiau această informație, nervul optic trimite semnalul către tectum, iar acesta activează mușchii. Drept urmare, limba țintește gângania cu o acuratețe impresionantă.

Logica acestui dispozitiv de tip imput-output a fost demonstrată grăitor de specialistul în neuroștiințe Roger Sperry. În anii 1960 a operat o broască – i-a scos ochii, i-a întors cu susul în jos și i-a reinserat.[0] Ochii funcționau. Broaștele au o capacitate uluitoare de regenerare. Nervul optic a crescut din nou de la ochi la tectum, restabilind harta vizuală internă. După ce broasca s-a vindecat și a putut vedea din nou, când o muscă zbura pe deasupra ei, limba îi țâșnea în jos. Dacă musca zbura la dreapta, limba țâșnea la stânga. Inteligența centrală a broaștei este o

---

[0] D. Ingle, „Visuomotor Functions of the Frog Optic Tectum”, Brain, Behavior, and Evolution 3 (1970): 57 – 71.

[0] R.W., „Sperryeffect of 180 Degree Rotation of the Retinal Field on Visuomotor Coordination”, Journal of Experimental Zoology Part A: Ecological and Integrative Physiology 92 (1943): 263 – 279.

mașinărie simplă, minunat de eficientă, care preia imputuri specifice ale nervilor și le conectează la outputurile corespunzătoare. Din nefericire, a fost păcălită prin manipulare științifică. Broasca modificată trebuia să fie hrănită de un om ca să nu moară de foame.

Tectumul broaștei nu se limitează la văz. Colectează și informații auditive de la receptorii tactili de la nivelul pielii.[0] Harta suprafeței corpului broaștei, a spațiului auditiv și vizual din jur converg și sunt integrate parțial în tectum. Acesta este cel mai înalt nivel de integrare al creierului amfibienilor – procesul central care adună semnalele disparate pe care le preia din mediu, selectează cel mai important eveniment care se întâmplă într-un anumit moment și declanșează o reacție.[0] Tectumul este mecanismul central al atenției broaștei.

Oamenii de știință pot sonda creierul cu o precizie uimitoare, la fel cum un specialist în computere sondează un circuit. O metodă standard implică un electrod – o sârmă dură, subțire cât un fir de păr, învelită în izolație de plastic, cu un capăt neacoperit. Doar o zecime de milimetru de fir este expusă. Ca un detector în miniatură, captează activitatea electrică la o distanță microscopică de metalul expus. Un cablu lung și flexibil conectat la partea din spate a electrodului ajunge la dispozitivele de monitorizare. Electrodul este de obicei fixat de mecanisme de precizie și este mutat cu câte un micrometru pentru a studia o anumită regiune cerebrală.

Această metodă este îndeajuns de sensibilă pentru a măsura activitatea neuronilor individuali din creier. Când un neuron aflat lângă capătul electrodului trimite un semnal către neuronii din apropiere, dispozitivul detectează acest impuls electric minuscul. Semnalul este amplificat și

---

[0] C. Comer și P. Grobstein, „Organization of Senzory Imputs to the Midbrain of the Frog, Rana pipiens”, Journal of Comparative Physiology 142 (1981): 161 – 168.

[0] B.E. Stein și M.A. Meredith, The Merging of the Senses (Cambridge, MA: MIT Press, 1993).

transmis către un difuzor, și oamenii de știință aud un clic. În mod normal, un neuron poate provoca un clic sau două pe secundă într-un tipar aleatoriu, dar, dacă apare un eveniment care activează neuronul, această celulă poate transmite rafale bruște de aproximativ o sută de clicuri pe secundă. Una dintre activitățile preferate ale specialiștilor în neuroștiințe este să asculte clicurile unui neuron individual și să se întrebe ce rol joacă în creier.

Fiecare neuron din tectumul broaștei funcționează ca un detector.[0] Neuronul monitorizează o anumită parte din spațiu – de exemplu, zona de deasupra capului –, clicurile intensificându-se când un obiect intră în acest spațiu. Neuronii sunt diferiți – unii preferă un stimul vizual care se mișcă într-un anumit fel, alții preferă o atingere sau un sunet. Cel puțin o parte dintre neuroni sunt multisenzoriali. Pentru un neuron multisenzorial nu contează dacă miști un obiect vizibil spre partea de sus a capului, dacă produci un sunet în același loc sau dacă stângi lumina și îți atingi regiunea superioară a capului; neuronul se activează și transmite semnalul celorlalte părți ale creierului. Dacă două sau toate trei simțurile converg și transmit același mesaj privind un obiect din apropiere, atunci neuronii din tectum se activează și mai intens. Se pare că un simplu calcul transmite mesajul: „Avem o informație, până aici totul e bine. Avem două sau trei care spun același lucru – pesemne că se petrece ceva important!”[0]

Aceeași metodă experimentală de bază poate fi aplicată invers, trimițând semnale electrice prin electrod pentru a activa neuronii – această metodă se numește microstimulare. Stimularea este atât de slabă încât nu s-ar simți pe piele, dar este îndeajuns de intensă pentru a activa neuronii și a-i face să genereze propriile semnale.

---

[0] C. Comer și P. Grobstein, „Organization of Senzory Imputs to the Midbrain of the Frog, Rana pipiens”, Journal of Comparative Physiology 142 (1981): 161 – 168; D. Ingle, „Visuomotor Functions of the Frog Optic Tectum”, Brain, Behavior, and Evolution 3 (1970): 57 – 71.

[0] B.E. Stein și M.A. Meredith, The Merging of the Senses (Cambridge, MA: MIT Press, 1993).

Microstimularea ne permite să ne punem întrebarea: dacă această mică grupare de neuroni de lângă capătul electrodului este activată artificial, ce anume îi spune animalului să facă?

Dacă este stimulat electric tectumul unei salamandre, animalul efectuează mișcări complicate și coordonate.[0] Se întoarce, își deschide gura, scoate limba, își întinde picioarele din față și își strânge ghearele ca și cum ar înhăța prada. Oricare ar fi spațiul monitorizat de neuronii dintr-un loc anume din tectum, stimularea electrică a acestora face ca animalul să se îndrepte într-acolo.

În cazul unei iguane, stimularea neuronilor din tectum specializați în cartografierea spațiului o face să-și întoarcă corpul, capul și ochii.[0] Animalul se uită direct spre punctul care reprezintă locul din harta mintală.

Dacă tectumul unui pește este stimulat într-un anumit loc, se va întoarce spre punctul spațial care corespunde acelui loc. În cazul ăsta, întoarcerea nu este la fel de simplă cum ar fi la alt animal întoarcerea capului, dat fiind că implică o interacțiune complexă între aripioare și apa din jur.

Șerpii din familia Crotalinae, cum ar fi șerpii cu clopoței și mocasinii de apă, au propria versiune de vedere în infraroșu: o pereche de organe specializate, sensibile la căldură, situate aproximativ la jumătatea distanței dintre ochi și nări. Aceste organe transmit informația către tectum, care conține o hartă a semnalelor de căldură suprapusă hărții vizuale a spațiului.[0] Capacitatea șarpelui

---

[0] T. Finkenstadt și J.—P. Ewert, „Visual Pattern Discrimination Through Interactions of Neural Networks: A Combined Electrical Brain Stimulation, Brain Lesion, and Extracellular Recording Study în Salamandra salamandra”, Journal of Comparative Physiology 153 (1983): 99 – 110.

[0] B.E. Stein și N.S. Gaither, „Senzory Representation în Reptilian Optic Tectum: Some Comparisons with Mammals”, Journal of Comparative Neurology 202 (1981): 69 – 87.

[0] P.H. Hartline, L. Kass și M.S. Loop, „Merging of Modalities în the Optic Tectum: Infrared and Visual Integration în Rattlesnakes”, Science 199 (1978): 1225 – 1229.

de a-și întoarce capul spre pradă și acuratețea loviturii depind de această hartă multisenzorială.

Tectumul bufniței are o hartă vizuală combinată cu o hartă auditivă.[0] Când vânează, poate direcționa lovitura fie în funcție de văz, fie, când vânează noaptea, în funcție de zgomotul pe care prada îl face în iarbă.

Stimularea colicului superior al unei maimuțe provoacă o mișcare bruscă și coordonată a capului și ochilor.[0] Maimuța se orientează spre locul corespunzător din spațiu. Nu știu dacă au fost realizate studii de stimulare electrică a colicului superior al oamenilor, dar suntem o specie de primate și se presupune că și la noi este prezent același mecanism ca la maimuțe. Când te întorci să te uiți la ceva, mai ales dacă un eveniment neașteptat te face să te întorci repede, din reflex, probabil că tectumul tău declanșează acest comportament bine coordonat, realizat aparent fără efort.

Toate vertebratele folosesc tectumul cam la fel, deși la unele lucrurile sunt mai complicate, în funcție de specie. Această arie cerebrală colectează informațiile senzoriale, selectează evenimentul cel mai pregnant care se desfășoară în apropiere și determină animalul să se uite în acea direcție, orientând fizic organele de simț spre acel eveniment.

Acest tip de orientare se numește uneori „atenție fățișă”.[0] Este o modalitate simplă de a rezolva o problemă fundamentală: în mediu se întâmplă prea multe lucruri deodată pentru a putea fi procesate de creier. E nevoie ca animalul să aleagă un obiect de interes și să le excludă pe celelalte. Dacă îți îndrepți ochii și urechile spre o țintă anume, atunci excluzi automat evenimentele periferice.

[0] S.P. Mysore și E.I. Knudsen, „The Role of a Midbrain Network în Competitive Stimulus Selection”, Current Opinion în Neurobiology 21 (2011): 653 – 660.

[0] R.H. Wurtz și J.E. Albano, „Visual-Motor Function of the Primate Superior Colliculus”, Annual Review of Neuroscience 3 (1980): 189 – 226.

[0] M.I. Posner, „Orienting of Attention”, Quarterly Journal of Experimental Psychology 32 (1980): 3 – 25.

Tectumul se ocupă de asta. Este primul centru de control al atenției care a apărut în creier pe parcursul evoluției.

De cele mai multe ori, când oamenii folosesc cuvântul *atenție*, se referă la atenția fățișă. În acest sens familiar al termenului, ești atent la ceea ce te uiți. Când nu te mai uiți la acel obiect sau dacă te întorci cu spatele, nu mai ești atent la el.

E limpede însă că privirea este doar o parte din povestea atenției. Un elev poate să mâzgălească o bucată de hârtie de pe bancă și în același timp să-și îndrepte atenția *mascat* către profesor. Sau să presupunem că tragi cu urechea la doi oameni care vorbesc despre tine. Nu vrei să te întorci și să te uiți la ei, dar atenția ta, resursele tale de procesare, se concentrează pe ascuns asupra acelei conversații. Sau poți să stai pe scaun pierdut în reverie și să fii atent la ceva ce nici măcar nu există în lumea fizică, în timp ce ochii tăi privesc în gol la tavan. În toate aceste exemple, atenția ta nu este îndreptată spre locul în care te uiți. Acest tip mai complex, mascat, de atenție nu face parte din atribuțiile tectumului, care nu se ocupă decât de atenția fățișă. Tectumul fiind în cazul broaștei principalul centru al atenției, aceasta nu are la dispoziție decât atenția fățișă, nu și pe cea mascată. Nu poate decât să se orienteze fizic spre obiectele din lumea înconjurătoare.

E inutil să ai atenție, orice fel de atenție, fie fățișă, fie mascată, dacă nu o poți controla. Dar controlul nu este o problemă simplă. Trebuie să monitorizezi îndeaproape elementul pe care îl controlezi. Pentru prima dată în povestea evoluției întâlnim nu doar celule care pot procesa informația și nu doar animale care-și pot direcționa atenția, ci sisteme cerebrale care construiesc o *schemă de atenție*- un pachet de informații, numit model intern, care monitorizează atenția. Povestea evoluției noastre se apropie de ceva ce aduce a conștiință – fără să fie întocmai asta.

O mașină care se conduce singură are nevoie de un model intern al mașinii. Nu este suficient ca informațiile

despre lumea din jur să ajungă la computerul mașinii și acesta să transmită apoi semnale către volan și pedale. Sistemul are nevoie de o serie de informații despre mașină, care să includă dimensiunea și forma, modul în care merge pe drum și starea sa, aflată în continuă schimbare – viteza, accelerația, poziția. Fără un model intern bogat, actualizat permanent, care să includă această gamă variată de informații, mașina ar avea un centru de control și ar putea transmite comenzi, dar ar face probabil un accident.

Acest principiu al unui model intern a fost aplicat prima oară în domeniul tehnic.[0] Nu contează ce este controlat – poate fi ceva fizic, precum o mașină sau brațul unui robot, sau poate fi ceva amorf, precum un curent de aer în încăperile dintr-o clădire mare. Pentru a funcționa corect, sistemul de control are nevoie de un model intern a ceea ce are de controlat. Are nevoie de o modalitate de a monitoriza mașina, sau robotul, sau curentul de aer. Sub unele aspecte, modelul intern este asemenea unei machete de pe masa generalului, cu tancuri și soldați mici de plastic. Este un set coerent de informații care, de obicei într-o modalitate simplificată și schematizată, reflectă și urmărește elementul care trebuie controlat.

Același principiu se aplică și la nivel biologic. Creierul controlează corpul cu ajutorul unui model intern, așa-numita schemă corporală, o serie de informații despre structura și starea în continuă schimbare a corpului.[0]

---

[0] E.F. Camacho și C. Bordons Alba, Model Predictive Control (New York: Springer, 2004); R.C. Conant și W.R. Ashby, „Every Good Regulator of a Sistem Must Be a Model of That Sistem”, International Journal of Sistems Science 1 (1970): 89 – 97; B.A. Francis și W.M. Wonham, „The Internal Model Principle of Control Theory”, Automatica 12 (1976): 457 – 465.

[0] M.S.A. Graziano și M.M. Botvinick, „How the Brain Reprezents the Body: Insights From Neurophysiology and Psychology”, în W. Prinz și B. Hommel (eds.), Common Mechanisms în Perception and Action: Attention and Performance XIX (Oxford, UK: Oxford University Press, 2002), 136 – 157; N. Holmes și C. Spence, „The Body Schema and the Multisensory Representation (s) of Personal Space”, Cognitive Processing 5 (2004): 94 – 105; F. De Vignemont, Mind the Body: An Exploration of Bodily Self-Awareness (Oxford, UK: Oxford University

Uneori, ca urmare a unui atac cerebral, pot apărea leziuni ale regiunilor cerebrale care structurează schema corpului.[0] Dacă un pacient nu mai știe ce formă sau structură are propria mână, nu va mai putea să o miște corespunzător; nu va mai putea îndeplini corect sarcini simple, cum ar fi să arate cu degetul, să întindă mâna sau să țină o cană. Dar nu trebuie să ne ducem atât de departe pentru a vedea rolul pe care îl joacă un model intern. Dacă îți agăți de încheietura mâinii o pungă grea de cumpărături și apoi încerci să apeși clanța unei uși, vei observa că la început mișcările tale sunt stângace. Modelul intern al mâinii pe care îl are creierul tău este dintr-odată greșit, dat fiind că dinamica fizică a membrului s-a schimbat. Foarte rapid, după doar câteva încercări, modelul tău intern începe să învețe noile reguli și mișcările tale devin mai line și mai precise.[0]

Dintr-o perspectivă tehnică, un model intern ar trebui să monitorizeze prezentul *și* să facă predicții despre viitor. Dacă vrei să controlezi ceva, cum ar fi un cărucior de cumpărături pe care îl împingi printre rafturi, ai nevoie să prezici ce e probabil să se întâmple în momentul următor. Creezi un fel de copie intuitivă a căruciorului, desfășori mișcările acestuia în minte și descoperi cum se comportă. Modul în care direcționezi căruciorul adevărat – forța pe care o aplici asupra mânerului – va depinde de predicțiile făcute de acel model intern. Copiii nu se pricep deloc la asta și lovesc căruciorul de rafturi, în parte pentru că nu au învățat să creeze un bun model intern al acestuia. Ei nu pot să prezică modul în care forța aplicată pe mâner va afecta mișcarea roților. Adulții, printr-o practică suficientă, își ajustează modelul inconștient, intuitiv.

Dar cum stau lucrurile cu atenția? Este, probabil, cel mai

Press, 2018).

[0] H. Head și G. Holmes, „Senzory Disturbances From Cerebral Lesions”, Brain 34 (1911): 102 – 254; G. Vallar și R. Ronchi, „Somatoparaphrenia: A Body Delusion. A Review of the Neuropsychological Literature”, Experimental Brain Research 192 (2009): 533 – 551.

[0] A.M. Haith și J.W. Krakauer, „Model-Based and Model-Free Mechanisms of Human Motor Learning”, în M. Richardson, M. Riley.

important proces al creierului și trebuie, fără îndoială, controlat. Pentru a răspunde eficient lumii, creierul trebuie să se concentreze conform unei strategii adecvate pe un obiect sau altul. Și, cu toate astea, atenția poate fi la fel de nesigură și capricioasă ca roțile unui cărucior de cumpărături și o poate lua ușor pe căi neașteptate. Conform principiilor de bază ale tehnicii controlului, știm că tectumul trebuie să se folosească de un fel de model intern pentru a urmări atenția. Colegii mei și cu mine am numit acest model intern ipotetic „schemă a atenției", paralelă cu schema corporală care contribuie la monitorizarea corpului. O schemă a atenției este un pachet de informații care descriu atenția – nu obiectul pe care îl vizează aceasta, ci atenția însăși. Monitorizează starea atenției, urmărește modul în care se poate schimba dinamic de la o stare la alta și prezice modalitatea în care se poate schimba în următoarele câteva momente. O versiune a unei scheme a atenției – informația care monitorizează specific atenția fățișă – a fost descoperită în tectumul maimuțelor și al pisicilor.[0] În principiu, același tip de informație este aproape sigur prezent la broaște, la pești și la orice alt animal care are tectum, chiar dacă nu a fost studiat explicit.

Să ne întoarcem acum la umila broască. Știm că are un procesor central, tectumul. Știm că are atenție fățișă, capacitatea de a-și îndrepta organele de simț spre un anumit punct din vastul mediu înconjurător. Știm că trebuie să aibă o schemă a atenției, dat fiind că e inutil să ai

[0] Există o amplă literatură despre experimentele asupra coliculului superior al pisicilor și maimuțelor, inclusiv asupra modului în care monitorizează și face predicții despre poziția capului și a ochilor și despre cum, în consecință, prezice modul în care imaginile vizuale vor trece prin retină. Menționez aici doar câteva analize. M.A. Basso și P.J. May, „Circuits for Action and Cognition: A View from the Superior Colliculus", Annual Review of Vision Science 3 (2017): 197 – 226; D.L. Sparks, „Conceptual issues related to the role of the superior colliculus în the control of gaze", Current Opinion în Neurobiology 9 (1999): 698 – 707; R.H. Wurtz și J.E. Albano, „Visual-Motor Function of the Primate Superior Colliculus", Annual Review of Neuroscience 3 (1980): 189 – 226.

atenție dacă nu poți s-o controlezi, și că nu poți s-o controlezi fără un model intern. Schema atenției este în fond un model sofisticat al sinelui. Nu doar că broasca își îndreaptă atenția spre obiecte specifice din lume, ci, într-un anumit sens, știe că face asta. Are informații *despre* propria atenție.

Ce anume știe creierul broaștei despre sine în virtutea faptului că are o schemă a atenției?

Să revenim la experimentul despre care am vorbit în Capitolul 2. Să luăm aparatul nostru futurist care transformă informația în limbaj, Vorbinator 5.000, și să-l conectăm la tectumul broaștei. Pe baza a ceea ce știm datorită schemei atenției, Vorbinatorul ar putea spune: „Aici sunt niște ochi. Aici e un corp. Ochii se mișcă încolo și-ncoace și se uită în diverse direcții. În momentul ăsta se uită la un punct negru care se mișcă în zigzag. Pentru că acum se mișcă, ochii vor ajunge în curând să se îndrepte în direcția aceea". Informația este atât de concretă pentru că broasca are o atenție limitată. Da, broasca are o schemă a atenției, dar este o schemă care nu se referă decât la atenția fățișă. Pentru broască, atenția înseamnă orientarea capului și a ochilor. Modelul intern de care are nevoie este așadar un model al capului și ochilor, al modului în care acestea se mișcă și se îndreaptă spre obiecte.

Să presupunem că, folosind Vorbinatorul, întrebăm tectumul broaștei: „Dar ai o *experiență* subiectivă privind acea muscă?"

Tectumul nu poate raporta decât informațiile pe care le deține. Spune:

— Acolo e un punct care se mișcă în zigzag. Aici sunt ochi. Aici e un corp. Ochii se mișcă. Se uită într-acolo.

Oarecum frustrați, spunem:

— Da, am înțeles. Dar ai *conștiință*? *Ai o reprezentare mintală* a acelei muște?

Tectumul broaștei răspunde:

— Aici sunt ochi. Aici e un corp. Ochii se uită într-acolo.

Tectumul broaștei pur și simplu nu deține informații pentru a putea răspunde la întrebările pe care i le adresăm.

Conceptul de conștiință este irelevant pentru el. În ciuda faptului că are un creier sofisticat, un tip de atenție și o schemă a atenției, broasca nu are nevoie de modele interne care să o reprezinte ca agent conștient.

Îmi sunt dragi în continuare Elvis și Priscilla. Știu că au un comportament minunat de bogat, din care face parte și cântecul lor de împerechere. Dacă aș petrece suficient timp în compania broaștelor, sunt sigur că aș dezvolta o relație cu ele și, lucru atât de caracteristic nouă, oamenilor, dat fiind că suntem creaturi sociale, aș ajunge chiar să intuiesc că micile animale au pe undeva o brumă de conștiință. Acestea sunt motivele umane, sociale pentru care oamenii ar putea avea senzația că o broască este conștientă. Este însă indiscutabil că broaștei îi lipsesc mijloacele prin care să creeze un model al conștiinței sau prin care să-și atribuie această proprietate. Poate să aibă cunoașterea *obiectivă* a propriei existențe și a mediului din jur, în sensul că procesează informațiile despre propriul corp și despre lume, dar, dacă am putea transforma aceste informații interne în limbaj, nu ar avea niciun motiv să susțină că are conștiință *subiectivă*.

Și, cu toate astea, elementele necesare sunt aproape la locul lor. Am văzut în această expunere evoluționistă că, acum circa jumătate de miliard de ani, primii pești fără maxilar inferior au dezvoltat o formă de atenție fățișă, un tectum care să o controleze și probabil o schemă a atenției care să faciliteze acest control. Amfibienii, reptilele, păsările și mamiferele, toate au moștenit acest sistem. Cu toții îl avem undeva adânc în noi. Dar pentru a descoperi ceva ce putem numi conștiință trebuie să mai facem un pas. Trebuie să trecem de la atenția fățișă la capacitatea mai complexă și mai subtilă a atenției mascate, la care păsările și mamiferele sunt experte.

4
Cortexul cerebral și conștiința

Cortexul cerebral, scoarța ridată atât de caracteristică creierului mamiferelor, a început să evolueze la reptile cu, probabil, peste 300 de milioane de ani în urmă, în perioada carboniferă[0] – epocă a marilor jungle mlăștinoase, când lumea era încă unită în supercontinentul Pangeea.[0] Volumul de oxigen din aer era mai mare decât astăzi și, drept urmare, insectele ajungeau la dimensiuni enorme, în ciuda faptului că aveau un sistem respirator ineficient.[0] Au fost găsite fosile de miriapode de 2, 5 metri și libelule de o jumătate de metru. Această perioadă este numită carboniferă pentru că, atunci când junglele luxuriante au dispărut în cele din urmă, cantitățile enorme de biomasă s-au transformat în zăcămintele de cărbune pe care le avem astăzi.

În prima parte a erei carbonifere lumea era plină de nevertebrate și amfibieni, singurele animale care se adaptaseră până atunci vieții pe uscat.[0] Reptilele au evoluat din amfibieni spre sfârșitul acestei ere, când climatul a început să fie uscat.[0] Au dezvoltat o piele impermeabilă, solzoasă, și ouăle pe care le depuneau au început să aibă coaja tare, rezistentă în afara apei.

Au dezvoltat și o parte a creierului care uneori este denumită *wulst*, o protuberanță a regiunii frontale anterioare. Această formațiune este la reptile o structură cu

---

[0] L. Medina și A. Reiner, „De Birds Possess Homologues of Mammalian Primary Visual, Somatosensory and Motor Cortices?”, Trends în Neurosciences 23 (2000): 1-12; R.K. Naumann și G. Laurent, „Function and Evolution of the Reptilian Cerebral Cortex”, în J. Kaas (coord.), Evolution of Nervous Sistems (San Diego: Academic Press, 2017), 491 – 518.

[0] R.R. Lemon, Vanished Worids: An Introduction to Historical Geology (Dubuque, IA: William C. Brown, 1993).

[0] J.F. Harrison, A. Kaiser și J.M. Vandenbrooks, „Atmospheric Oxygen Level and the Evolution of Insect Body Size”, Proceedings: Biological Science 277 (2010): 1937 – 1946.

[0] R.L. Carroll, „The Origin and Early Radiation of Terrestrial Vertebrates”, Journal of Paleontology 75 (2001): 1202 – 1213.

[0] S. Sahney, MJ. Benton și HJ. Falcon-Lang, „Rainforest Collapse Triggered Pennsylvanian Tetrapod Diversification în Euramerica”, Geology 38 (2010): 1079 – 1082.

precădere senzorială. Conține o hartă organizată a spațiului vizual și o hartă a receptorilor tactili de la nivelul pielii, relativ asemănătoare cu hărțile senzoriale din tectum despre care am scris în capitolul anterior. *Wulstul* este un fel de tectum 2.0, o revoluție în termeni evoluționiști ce poate fi comparată cu trecerea de la telefonul fix la iphone.

Se crede pe nedrept despre reptile că ar fi proaste. Prin „creier reptilian” se înțelege în general partea străveche care zace sub zonele inteligente. Dar adevărul este că reptilele au deja începutul a ceea ce a devenit cortexul cerebral uman. Au o adevărată scânteie de inteligență. Multe specii de reptile au capacitatea de a rezolva probleme și realizează interacțiuni sociale complexe.[0]

La scurt timp după apariția reptilelor în era carboniferă, acestea s-au împărțit în două grupuri principale. Sinapsidele, așa-numite reptile care seamănă cu mamiferele, au fost inițial doar puțin diferite de sauropside, care formează celălalt grup major.[0] Sinapsidele aveau o structură a craniului ușor diferită, cu o motricitate mai bună a mandibulei. Simplu spus, se pricepeau mai bine să mănânce, ceea ce le oferea un avantaj din punct de vedere al energiei. Cel mai bine cunoscut exemplu de sinapsidă timpurie este *Dimetrodon*, din perioada permiană. E acea creatură preistorică din orice set de dinozauri de jucărie care pare că se târăște, seamănă cu un crocodil și are pe spate o creastă mare cu țepi. Dar nu e un dinozaur. E o reptilă asemănătoare mamiferelor, care seamănă mai mult

---

[0] M. Leal și B.J. Powell, „Behavioural Flexibility and Problem-Solving în a Tropical Lizard”, Biological Letters 8 (2012): 28 – 30; J.D. Manrod, R. Hartdegen și G.M. Burghardt, „Rapid Solving of a Problem Apparatus by Juvenile Black-Throated Monitor Lizards (Varanus albigularis albigularis)”, Animal Cognition 11 (2008): 267 – 273; R.T. Mason și M.R. Parker, „Social Behavior and Pheromonal Communication în Reptiles”, Journal of Comparative Physiology A: Neuroethology, Senzory, Neural, and Behavioral Physiology 196 (2010): 729 – 749.

[0] T.S. Kemp, The Origin and Evolution of Mammals (Oxford, UK: Oxford University Press, 2005); A.S. Römer și L.W. Price, „Review of the Pelycosauria”, Geological Society of America, Special Papers 28 (1940): 1 – 534.

cu noi decât cu *Tyrannosaurus rex* – și probabil stătea mai ridicată, aducând mai puțin cu o șopârlă care-și târăște burta pe pământ decât ai putea crede uitându-te la figurinele greoaie de plastic.

Pe măsură ce reptilele asemănătoare mamiferelor s-au transformat treptat în mamiferele moderne în decurs de sute de milioane de ani, *wulstul* s-a dezvoltat și a devenit o structură stratificată de neuroni care acoperea cea mai mare parte a creierului.[0] La multe mamifere, cortexul este un strat fin care învelește creierul. Dar la unele specii și-a extins atât de mult dimensiunea și a ajuns atât de înghesuit încât s-a mototolit ca o bucată de pânză îndesată în craniu. Cortexul omului, dacă ar fi întins, ar ajunge cam la dimensiunea și grosimea unui prosop mare de mâini.

S-au extins și alte structuri cerebrale care au o legătură apropiată cu cortexul – în special o formațiune mare, asemănătoare unui avocado, situată la baza creierului și numită talamus.[0] Denumirea este potrivită. *Thalamus* înseamnă „pat” sau „fundație” în latină (inițial, camera în care se afla patul, în elină). Fiecare parte din cortex este legată de talamus și majoritatea informațiilor care ajung la cortex trec mai întâi prin talamus, ceea ce l-a făcut să-și câștige porecla de „poartă a cortexului”. Rigoarea mi-ar impune să folosesc termenul tehnic „sistem talamo-cortical”, dar, de dragul simplității, voi continua să folosesc prescurtarea „cortex” când vorbesc despre acest circuit specializat.

Pe măsură ce reptilele asemănătoare mamiferelor și-au dezvoltat cortexul sofisticat în decursul a 300 de milioane de ani, reptilele care nu semănau cu mamiferele, sauropsidele, au găsit o altă cale ca să-și sporească

---

[0] Z. Molnar, J.H. Kaas, J.A. De Carlos, R.F. Hevner, E. Lein și P. Nëmec, „Evolution and Development of the Mammalian Cerebral Cortex”, Brain, Behavior, and Evolution 83 (2014): 126 – 139.

[0] E.G. Jones, The Thalamus (New York: Springer, 1985); A.B. Butler, „Evolution of the Thalamus: A Morphological and Functional Review”, Thalamus and related sistems 4 (2008): 35 – 58.

inteligența. Această cale a dus la animalele numite archozauri, apoi la dinozauri și, în cele din urmă, la păsările din ziua de azi. Archozaurii erau reptile prădătoare mari, cu picioare scurte, și aveau un creier puțin mai mare decât predecesorii lor imediați.[0] Crocodilii, archozaurii din prezent, sunt unele dintre reptilele cele mai inteligente și au un comportament complex.[0] Se poate observa că sunt inteligenți din șiretenia cu care-și urmăresc prada și din modul în care împart hrana și își îngrijesc puii.

În urmă cu aproximativ 230 de milioane de ani, în perioada triasică, archozaurii au produs un subgrup specializat, o creatură bizară care a evoluat astfel încât să poată alerga pe picioarele alungite din spate.[0]13 Pentru aceia dintre voi care au în minte imaginea unor dinozauri giganți care mergeau pe patru picioare, fosilele ne arată că toți dinozaurii au evoluat dintr-un tip inițial, străvechi, care mergea în două picioare și, probabil, se sprijinea uneori cu mâinile pe pământ. Ulterior, unele specii de erbivore s-au întors la mersul în patru labe, în timp ce dinozaurii prădători au rămas bipezi.

Toată lumea cunoaște clișeul potrivit căruia dinozaurii ar fi avut creierul cât o nucă. Această afirmație este de-a dreptul scandaloasă. Craniile fosilizate ne arată ceva despre mărimea și structura creierului lor.[0] Cei mai mari

---

[0] P. Senter, „Phylogenetic Taxonomy and the Names of the Major Archosaurian (Reptilia) Clades”, Paleobios 25 (2005): 1 – 7.

[0] V. Dinets, „Aparent Coordination and Collaboration în Cooperatively Hunting Crocodilians”, Ethology, Ecology, and Evolution 27 (2012): 244 – 250; J.S. Doody, G.M. Burghardt și V. Dinets, „Breaking the Social-Non-Social Dihotomy: A Role for Reptiles în Vertebrate Social Behavior Research?”, Ethology 119 (2012): 1 – 9; L.D. Garrick și J.W. Lang, „Social Signals and Behaviors of Adult Alligators and Crocodiles”, American Zoologist 17 (1977): 225 – 239.

[0] M.C. Langer, M.D. Ezcurra, J.S. Bittencourt și F. Novas, „The Origin and Early Evolution of Dinosaurs”, Biological Reviews 85 (2010): 55 – 110.

[0] M. Bronzați, O.W.M. Rauhut, J.S. Bittencourt și M.C. Langer, „Endocast of the Late Triassic (Camian) Dinozaur Saturnalia tupiniquim. Implications for the Evolution of Brain Tissue în Sauropodomorpha”, Scientific Reports 7 (2017): 11931; S.W. Rogers, „Allosaurus, Crocodiles, and Birds: Evolutionary Clues From Spiral Computed

dinozauri aveau un creier proporțional cu masa lor totală. Creierul unui *T. Rex*, de exemplu, rivaliza cu cel omenesc ca dimensiune, deși avea în mod sigur mult mai puțini neuroni și conexiuni.[0] Dinozaurii aveau aceeași organizare cerebrală de bază precum crocodilii, inclusiv acel protocortex, *wulstul*. Nu erau proști. În special teropodele – dinozaurii prădători, bipezi –, care erau probabil cele mai inteligente animale de pe planetă în acea perioadă. Având ochii în față și vedere stereoscopică, probabil și-au dezvoltat partea vizuală a *wulst-ului* pentru a putea procesa mai bine marea cantitate de informații absorbite.

Păsările, așa cum știe orice școlar, au evoluat din dinozauri. Dar această informație este o simplificare a unui adevăr mai subtil și mai ciudat. Zborul păsărilor pare să fi apărut treptat de-a lungul jurasicului, undeva între 200 și 145 de milioane de ani în urmă.[0] Nu e de niciun folos să indicăm un moment anume și să spunem: „Uite, atunci au apărut păsările”. Lipsa unei delimitări clare între dinozauri și păsări apare limpede mai ales din dovezile oferite de fosilele de dinozauri cu pene păstrate în cenușă vulcanică și descoperite în stratul fosilifer din provincia Liaoning din China.[0] Adevărul care reiese acum este că păsările *sunt* dinozauri teropozi. Dacă ai fi trăit în mezozoic, nu te-ai fi gândit nicio clipă că păsările erau o categorie separată.

Tomography of an Endocast”, Anatomical Record 2 57 (1999): 162 – 173.

[0] L.M. Witmer și R.C. Ridgely, „New Insights Into the Brain, Braincase, and Ear Region of Tyrannosaurs (Dinosauria, Theropoda), With Implications for Senzory Organization and Behavior”, Anatomical Record 292 (2009): 1266 – 1296.

[0] S.L. Brusatte, J.K. O’Connor și E.D. Jarvis, „The Origin and Diversification of Birds”, Current Biology 25 (2015): R888-R898; L.M. Chiappe, Glorified Dinosaurs: The Origin and Early Evolution of Birds (Hoboken, NJ: John Wiley and Sons, 2007); Z. Zhou, „The Origin and Early Evolution of Birds: Discoveries, Disputes, and Perspectives From Fossil Evidence”, Naturwissenschaften 91 (2004): 455 – 471.

[0] Q. Ji și S. Ji, „On the Discovery of the Earliest Bird Fossil în China (Sinosauropteryx) And The Origin of Birds”, Chinese Geology 10 (1996): 30 – 33; M.A. Norell și X. Xu, „Feathered Dinosaurs”, Annual Review of Earth and Planetary Science 33 (2005): 277 – 299.

Lumea era plină de dinozauri cu pene, unii mari, alții mici, unii cu dinți, alții cu cioc, care mergeau țanțoș pe două picioare, săltând din cap, sau care zburau. Dacă ai fi văzut o găină alergând sau o vrabie zburând, te-ai fi gândit: „Ia te uită! încă o variantă a dinozaurului teropod!” în mod cert nu ai fi spus: „Iată, o nouă clasă de animale care a evoluat din dinozauri!” La fel, nimeni n-ar spune astăzi că liliecii au evoluat *din* mamifere. Liliecii *sunt* mamifere care s-au adaptat zborului. Astfel, păsările sunt dinozaurii zburători care au supraviețuit până în prezent. Iar specialiștii în neuroștiințe le-au studiat îndelung.

Păsările au un *wulst* cu mult mai mare decât cel al reptilelor.[0] Timp de sute de milioane de ani categoria mamiferelor și-a ajustat creierul, dezvoltând *wuls*tul reptilian într-un cortex elaborat; categoria dinozaurilor/păsărilor a evoluat la rândul ei, găsindu-și propriile moduri de a schimba și de a elabora aceeași structură cerebrală.

Deși expresia „minte de găină” încă este considerată a fi o insultă, păsările sunt animale inteligente. Unele specii au vieți sociale complexe, altele sunt vânători șireți, iar altele au o excelentă memorie a rezervelor de hrană. În special ciorile sunt renumite pentru inteligența lor cognitivă. Pot să îndoaie sârme pentru a le transforma în cârlige, pe care le folosesc ca unelte cu care să ajungă la hrană.[0]19 Celebra fabulă a lui Esop în care cioara aruncă

---

[0] H.J. Karten, „Vertebrate Brains and Evolutionary Connectomics: On the Origins of the Mammalian ’Neocortex”, Philosophical Transactions of the Royal Society of London, B, Biological Sciences 370 (2015): 20150060;
L. Medina și A. Reiner, „De Birds Possess Homologues of Mammalian Primary Visual, Somatosensory and Motor Cortices?”, Trends în Neurosciences 23 (2000): 1-12.

[0] G.R. Hunt și R.D. Gray, „Tool Manufacture by New Caledonian Crows: Chipping Away at Human Uniqueness”, Acta Zoologica Sinica (Supplement) 52 (2006): 622 – 625; C. Rutz și J.J. St Clair, „The Evolutionary Origins and Ecological Context of Tool Use în New Caledonian Crows”, Behavioral Processes 89 (2012): 153 – 165; C. Rutz, S. Sugasawa, J.E. Van der Wal, B.C. Klump și J.J. St Clair, „Tool Bending în New Caledonian Crows”, Royal Society of Open Science 3 (2016):

pietre în ulcior pentru a crește nivelul apei până când poate apuca bucățica de hrană care plutește nu este un mit. Ciorile pot să rezolve această sarcină.[0]

Inteligența remarcabilă a păsărilor depinde probabil de dezvoltarea regiunii anterioare a creierului lor. Este îndoielnic că dinozaurii dispăruți erau la fel de inteligenți cum sunt ciorile de azi – ciorile reprezintă un progres evoluționist recent –, dar aceleași structuri cerebrale bazale sunt prezente în întreaga descendență reptile-dinozauri-păsări.

Mă voi concentra asupra cortexului mamiferelor, pentru că despre acesta se cunosc mult mai multe și pentru că singurele animale de azi care pot atesta verbal faptul că sunt conștiente – oamenii – sunt mamifere. Dar este bine să nu uităm că alte tipuri de animale, ca păsările sau crocodilii, se bazează pe același sistem cerebral pentru comportamentul lor complex și variat.

Acum, că am făcut o scurtă prezentare a evoluției cortexului cerebral, voi explica de ce cred eu că această structură susține conștiința. E necesar pentru asta să introducem unele detalii despre modul în care cortexul procesează informația. Nu este doar un procesor mai mare montat peste modele mai mici și mai vechi, cum ar fi tectumul din creierul broaștei. Este un tip fundamental diferit de procesor. Este o mașinărie calibrată să cearnă un volum enorm de informații și să le separe în subcategorii mici. Aceste subcategorii sunt procesate într-o manieră profundă, exhaustivă și ajung să ghideze comportamentul. Dat fiind acest proces constant de cernere și separare, cortexul este în mod fundamental o mașinărie a atenției.

Cortexul cerebral este ca Liga Națională de Fotbal (National Football League, NFL). În NFL, echipele trec prin

---

160439.

[0] S.A. Jelbert, A.H. Taylor, L.G. Cheke, N.S. Clayton și R.D. Gray, „Using the Aesop’s Fable Paradigm to Investigate Causal Understanding of Water Displacement by New Caledonian Crows”, PLOS One 9 (2014): e92895.

diferite runde eliminatorii, până când în finală, la Super Bowl, una iese câștigătoare. În cortex, informația trece prin niveluri diferite de procesare, aflându-se permanent în competiții intense, în care câștigătorul ia tot.[0] Indiferent de tipul de informație – vizuală, auditivă, afectivă, intelectuală –, arhitectura cortexului creează runde eliminatorii. Informația din ce în ce mai atent selectată este procesată din ce în ce mai în profunzime și are șanse din ce în ce mai mari să aibă un efect asupra comportamentului. În cele din urmă, un pachet integrat de informații câștigă finala corticală și devine elementul care, cel puțin pentru moment, se află în centrul procesării.

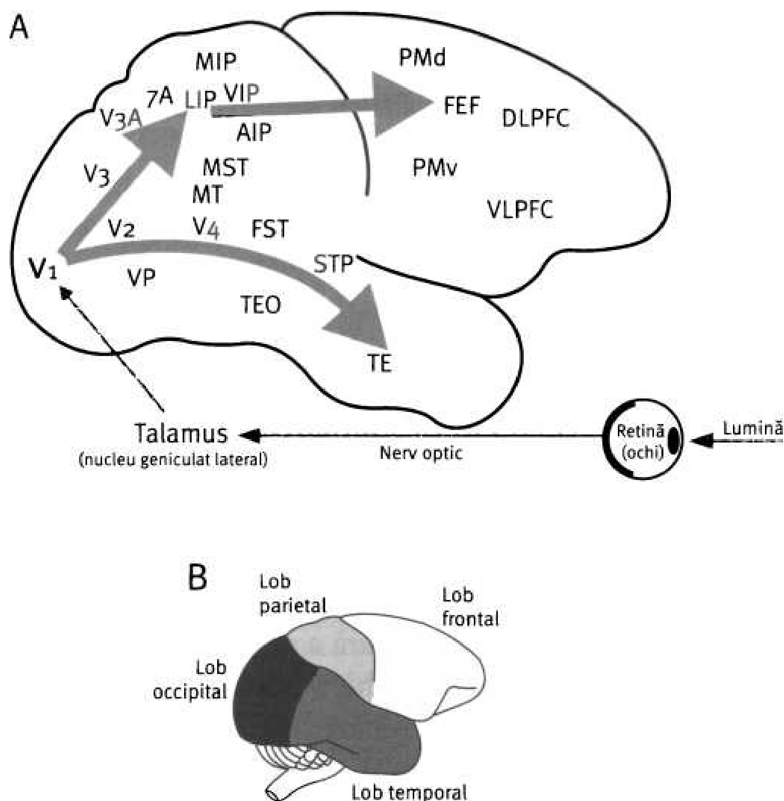
Acest stil eliminatoriu de procesare a fost studiat cel mai mult în cazul văzului. Figura 4.1 prezintă o schemă parțială a sistemului vizual, bazată pe zeci de ani de cercetări efectuate asupra primatelor, deopotrivă maimuțe și oameni.[0] Lumina pătrunde în ochi și formează o imagine pe retină – țesutul sensibil la lumină situat în partea din spate a ochiului. Neuronii din retină se activează și intră imediat într-o competiție locală. Semnalele concurează unele cu

---

[0] D.M. Beck și S. Kastner, „Top-Down and Bottom-Up Mechanisms în Biasing Competition în the Human Brain”, Vision Research 49 (2009): 1154 – 1165; R. Desimone și J. Duncan, „Neural Mechanisms of Selective Visual Attention”, Annual Review of Neuroscience 18 (1995): 193 – 222.

[0] Este enormă literatura despre mozaicul regiunilor corticale vizuale ale primatelor. Au contribuit la acest domeniu mii de oameni, printre care și eu. Menționez aici doar câteva surse de informare în cazul maimuțelor și al oamenilor. D. Felleman și D. Van Essen, „Distributed Hierarchical Processing în the Primate Visual Cortex”, Cerebral Cortex 1 (1991): 1 – 47; K. Grillspector și R. Malach, „The Human Visual Cortex”, Annual Review of Neuroscience 27 (2004): 649 – 677; P. Schiller și E. Tehovnik, Vision and the Visual Sistem (Oxford, UK: Oxford University Press, 2015); L.G. Ungerleider și J.V. Haxby, „ «What» and «Where» în the Human Brain”, Current Opinion în Neurobiology 4 (1994): 157 – 165; D.C. Van Essen, J.W. Lewis, H.A. Drury, N. Hadjikhani.
R. B. Tootell, M. Bakircioglu și M.I. Miller, „Mapping Visual Cortex în Monkeys and Humans Using Surface-Based Atlases”, Vision Research 41 (2001): 1359 – 1378; L. Wang, R.E. Mruczek, M.J. Arcaro și
S. Kastner, „Probabilistic Maps of Visual Topography în Human Cortex”, Cerebral Cortex 25 (2015): 3911 – 3931.

altele, cele mai puternice suprimându-le pe cele mai slabe. Semnalele care corespund contrastelor mari sunt favorizate.



4.1 Parte dintre regiunile corticale vizuale din creierul primatelor. (A) O subcategorie mică a principalelor arii implicate în văz și parcursul general al informației prin aceste regiuni. (B) Cei patru lobi ai cortexului cerebral.

Informațiile pornesc apoi din ochi prin nervul optic, un cablu de aproximativ 1,5 milioane de fibre, și ajung la talamus, situat la baza creierului. Informațiile vizuale ajung într-o zonă specifică a talamusului, o protuberanță situată pe o parte a acestuia numită nucleu geniculat lateral (semnificația literală în latină ar fi: „nucleul care se reliefează într-o parte și care arată de parcă ar fi genunchiul cuiva”). Din nou, în talamus informațiile

concurente sunt filtrate corespunzător.

După ce vor fi trecut prin nucleul geniculat lateral, informațiile călătoresc apoi de-a lungul fibrelor până la primul nivel al cortexului vizual, o regiune din lobul occipital numită cortex vizual primar sau V1, unde se desfășoară din nou același tip de competiție între neuroni.

De la V1, informațiile trec printr-o încrengătură de regiuni vizuale de ordin superior – V2, V3, V4, MT, MST, TEO, TE și așa mai departe –, un mișmaș de denumiri. În ultimii 50 de ani, specialiștii în neuroștiințe au cartografiat cortexul vizual din ce în ce mai amănunțit, identificând zeci de regiuni și subregiuni care acoperă aproximativ 60% din cortex, în special în lobii occipital, temporal și parietal. În fiecare dintre aceste regiuni de mărimea unui timbru poștal, pachete mici de informații sunt permanent în competiție, neuronii inhibându-se unii pe alții.

Ariile vizuale au o organizare generală, procesarea fiind din ce în ce mai complexă pe măsură ce informațiile trec din zonele inferioare de la baza creierului în zonele superioare, mai apropiate de partea anterioară a creierului. De exemplu, neuronii din V1, de la baza ierarhiei, împart lumea vizuală în segmente scurte și pete de culoare. Fac o împărțire simplă, superficială, dar detaliată a lumii vizuale. Prin opoziție, neuronii din TE, o regiune situată la un nivel superior al ierarhiei, sunt sensibili la imagini complexe, cum ar fi fețele și mâinile. Aceștia procesează informații despre identitatea obiectelor, nu despre detalii. Și totuși parcursul informațiilor nu se desfășoară exclusiv pe scara ierarhică. Semnalele pot să ajungă înapoi și în lateral, în orice direcție și în toate direcțiile din rețea. Pentru ca lucrurile să fie și mai complexe, cortexul comunică permanent cu regiunile cerebrale aflate mai în profunzime, în special cu talamusul.

Pe măsură ce informația trece prin acest sistem de procesare, competiția dintre semnale seamănă tot mai mult cu o versiune deformată a NFL. Semnalul cel mai puternic nu câștigă întotdeauna. Alte influențe pot să-și lase amprenta. Când te uiți în jur, s-ar putea ca un element să fie cel mai strălucitor sau cel mai rapid, generând cel mai

puternic semnal, dar nu domină în mod necesar procesarea cerebrală. Un semnal transmis de lobul frontal poate ajunge în cortexul vizual și poate determina procesarea unui alt element, mai puțin evident, făcându-l pe acesta să câștige competiția.[0] Când competiția este câștigată de un element care iese în evidență, acest proces este numit atenție de tip *bottom-up*, iar când un semnal intern își pune amprenta și schimbă centrul de interes, procesul este denumit atenție de tip *top-down*[0] Dar acestea sunt de fapt aspecte diferite ale aceluiași proces subiacent.

Specialistul în neuroștiințe Robert Desimone a contribuit la definirea acestei îmbulzeli competitive din cortex și i-a dat denumirea potrivită de „competiție părtinitoare" (*biased competition*).[0] Cred că acesta este unul dintre adevărurile organizării primare a cortexului. Inhibiția locală a neuronilor, care determină competiția, domină mașinăria cortexului. Nu întâmplător epilepsia este o boală a cortexului. Criza epileptică apare când inhibiția locală nu funcționează.[0] Semnalele care în mod normal se controlează unele pe altele proliferează dintr-odată și devin valuri nestăvilite de activitate care inundă cortexul. Această boală demonstrează faptul că inhibiția este esența

[0] T. Moore și M. Zimsak, „Neural Mechanisms of Selective Visual Attention", Annual Review în Psychology 68 (2017): 47 – 72.

[0] Bottom-up (literal, de jos în sus) desemnează tipul de procesare care pornește de la structurile senzoriale (simple) spre cele mai complexe, de interpretare a sensului. Top-down (de sus în jos) desemnează tiparul invers de procesare, de la structurile superioare (e.g. cognitive) la cele simple. De exemplu, procesul de identificare a sensului unui stimul tactil ca fiind o mângâiere sau o împunsătură este de tip bottom-up, la fel este și procesul de a atribui un cuvânt unui obiect din mediu - sunt operații care pornesc de la un input senzorial, simplu, și ajung la o procesare complexă. Un exemplu de procesare de tip top-down este operația de rotire mintală a unui obiect în spațiu, anterioară manipulării fizice a obiectului - este un tip de operație care se îndreaptă de la aspecte complexe spre cele simple (n. tr.).

[0] R. Desimone și J. Duncan, „Neural Mechanisms of Selective Visual Attention", Annual Review of Neuroscience 18 (1995): 193 – 222.

[0] G. Alarcón și A. Valentin, Introduction to Epilepsy (Cambridge, UK: Cambridge University Press, 2012).

cortexului și arată modul în care sistemul eșuează drastic fără o inhibiție suficientă.

În Capitolul 2 am descris un truc simplu utilizat de ochiul crabului, numit inhibiție laterală, care constă în blocarea reciprocă a neuronilor.[0] Rezultatul inhibiției laterale este intensificarea contrastului imaginii. Zonele luminoase sunt înregistrate ca fiind mai luminoase; zonele întunecate sunt înregistrate ca fiind și mai întunecate. Competiția părtinitoare este varianta de inhibiție laterală a cortexului, amplificată de un milion de ori, extinsă de la competiția locală de la nivelul ochiului la o lume ticsită de runde eliminatorii ordonate într-o ierarhie vastă.

Această ierarhie nu se limitează la văz. După ce este filtrată prin rețele interconectate din regiunile cerebrale, informația vizuală ajunge în cele din urmă în competiție cu alte tipuri de informație. Atingerea, auzul, controlul mișcărilor, gândirea abstractă – domenii vaste de informații converg în regiuni de ordinul cel mai înalt, în rețelele integratoare ale creierului. Aceste zone se găsesc în special în lobii parietal și frontal, fiind așadar numite, fără prea multă imaginație, rețele parietal-frontale.[0] Informația care

[0] R.B. Barlow Jr. Și A.J. Fraioli, „Inhibition în the Limulus Lateral Eye în Situ”, Journal of General Physiology 71 (1978): 699 – 720; K. Hadeler, „On the Theory of Lateral Inhibition”, Kybernetik 14 (1974): 161 – 165.

[0] M. Corbetta, G. Patel și G.L. Shulman, „The Reorienting Sistem of the Human Brain: From Environment to Theory of Mind”, Neuron 58 (2008): 306 – 324; K. Igelström și M.S.A. Graziano, „The Inferior Parietal Lobe and Temporoparietal Junction: A Network Perspective”, Neuropsychologia 105 (2017): 70 – 83; R. Saxe și L.J. Powell, „It’s the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind”, Psychological Science 17 (2006): 692 – 699; M. Școlari, K.N. Seidl-Rathkopf și S. Kastner, „Functions of the Human Frontoparietal Attention Network: Evidence From Neuroimaging”, Current Opinion în Behavioral Sciences 1 (2015): 32 – 39; f. L. Vincent, I. Kahn, A.Z. Snyder, M.E. Raichle și R.L. Buckner, „Evidence for a Frontoparietal Control Sistem Revealed by Intrinsic Functional Connectivity”, Journal of Neurophysiology 100 (2008): 3328 – 3342; B.T.T. Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, D. Lashkari, M. Hollinshead, J.L. Roffman, J.W. Smoller, L. Zollei, J.R. Polimeni, B. Fischl, H. Liu și R.L. Buckner, „The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity”, Journal of

ajunge la aceste rețele a câștigat finala. A obținut ceea ce filosoful Daniel Dennett a numit *fame în the brain* (celebritate în creier). Mașinăria corticală de runde eliminatorii a redus lumea la câteva elemente care influențează acum rețelele centrale ale cortexului și au cele mai mari șanse să influențeze comportamentul și să intre în memorie. Însă ce este în atenție acum nu va rămâne neapărat așa în momentul următor. Competiția este tot timpul în schimbare, este permanent dinamică, noi informații câștigând lupta din rețelele neuronale și ajungând la cel mai înalt nivel de procesare. Sistemul cortical este neobosit prin însăși natura modului său de operare. Competiția este intrinsec instabilă.

Acest proces de cernere neobosită este cel mai puternic tip de atenție pe care îl avem. E cel mai sofisticat instrument de care creierul se folosește pentru a înțelege lumea. Sunt uneori întrebat de ce am construit o teorie a conștiinței pornind de la un subiect atât de specific, aparent limitat, ca atenția. Dar atenția nu este un subiect limitat. Este esența modului în care cortexul cerebral înțelege lumea.

Tectumul, acea structură cerebrală primitivă, veche de o jumătate de miliard de ani, despre care am vorbit în capitolul anterior, coordonează atenția fățișă. Poate să determine direcția în care ne îndreptăm ochii și urechile, asemenea unei antene de satelit, pentru a asimila mai multe informații din acea parte a lumii. Evoluția cortexului a făcut posibilă însă o nouă abordare a procesării lumii: atenția mascată. Poți să fii atent la ceva ce nu se află în fața ochilor tăi.

Nu vreau să spun că cortexul nu are nimic de-a face cu mișcarea ochilor. Joacă un rol important în coordonarea ochilor și a capului.[0] Dar atenția corticală nu se

---

Neurophysiology 106 (2011): 1125 – 1165.

[0] C.L. Colby și M.E. Goldberg, „Space and Attention în Parietal Cortex”, Annual Review of Neuroscience 22 (1999): 319 – 349; J. Gottlieb, „From Thought to Action: The Parietal Cortex as a Bridge Between Perception,

concentrează neapărat în direcția în care sunt îndreptați ochii. Poți să-ți legi neliniștit șireturile în timp ce atenția corticală îți este îndreptată spre șeful care stă în fața ta. Poți să fii atent la o mâncărime de pe spate, ca și cum lumea întreagă s-ar fi redus la senzația aceea neplăcută, și în același timp să zâmbești și să te uiți în ochii prietenului tău fără să auzi nimic din ce-ți spune. Poți să te holbezi la o pagină dintr-o carte și atenția să ți se îndrepte spre o întâmplare petrecută cu o zi în urmă – un gând care nu are nicio legătură cu direcția în care privești.

Diferența dintre atenția fățișă și cea mascată este următoarea: prima înseamnă să cuprinzi un element cu organele de simț; cea de-a doua înseamnă să cuprinzi un element cu mașina de calcul masivă a cortexului. Poți să „îndrepți" această mașinărie spre orice, fie că e vorba de un obiect concret din fața ta, de ceva aflat la marginea câmpului tău vizual ori de ceva intern: un gând sau o emoție.

Oamenii de știință au comparat atenția mascată cu lumina unui reflector.[0] Conform acestei analogii, reflectorul luminează intens un obiect sau două și aruncă în jur o lumină difuză care poate face vizibile și alte lucruri. Dintre sutele de obiecte care sunt în orice moment în jurul tău, reflectorul cortical iluminează o parte, în timp ce restul obiectelor sunt total lăsate pe dinafară, complet neînregistrate. Această analogie este însă ușor forțată dacă luăm în considerare faptul că reflectorul nu se poate mișca decât în spațiul fizic, în timp ce atenția mascată poate aborda dimensiuni care n-au nicio legătură cu spațiul.[0]

Action, and Cognition", Neuron 53 (2007): 9 – 16; E.J. Tehovnik. M.A. Sommer, LH. Chou, W.M. Slocum și P.H. Schiller, „Eye Fields în the Frontal Lobes of Primates", Brain Research Reviews 32 (2000): 413 – 448.

[0] C. Eriksen și J. St fames, „Visual Attention Within and Around the Field of Focal Attention: A Zoom Lens Model", Perception and Psychophysics 40 (1986): 225 – 240; M.I. Posner, C.R. Snyder și B.J. Davidson, „Attention and the Detection of Signals", Journal of Experimental Psychology 109 (1980): 160 – 174.

[0] M. Școlari, E.F. Ester și J.T. Serences, „Featureand Obiect-Based

Atunci când privești un tablou de Mondrian, poți să-ți concentrezi atenția exclusiv asupra culorilor, ignorând total formele; drept urmare, nu-ți vei mai aminti deloc formele, chiar dacă erau prezente alături de culori în același spațiu. Acest tip de atenție nu prea seamănă cu un reflector real. Din aceste considerații, analogia cu reflectorul și-a pierdut din relevanță. Personal, îmi place analogia, câtă vreme reflectorul luminează într-un spațiu al dimensiunilor abstracte, și nu doar în cele trei dimensiuni ale spațiului fizic.

Cortexul cerebral creează acest reflector intern. Ne permite să explorăm un peisaj aproape infinit, multidimensional, în care putem să procesăm orice, de la obiectele cele mai concrete și mai apropiate la ideile cele mai abstracte.

Nu poți să procesezi în profunzime fără să schimbi centrul atenției. Ca să procesezi în profunzime toate informațiile disponibile simultan, ar trebui să ai un creier de dimensiunea unei planete. Evoluția a găsit un mod eficient de a utiliza un volum limitat de putere cerebrală pentru a procesa lumea într-un mod inteligent. Soluția a fost să supună informația unei competiții atât de strânse, încât doar o mică parte să fie procesată complet într-un anumit moment și să construiască în sistem un panou de comandă sofisticat pentru modificarea și ajustarea centrului atenției. Atenția este cheia care deschide înțelegerea complexă a lumii.

Cercetătorii care scriu despre evoluția conștiinței pun accentul pe creșterea treptată a gradului de complexitate a creierului. Este tentant să credem că această complexitate determină conștiința.[0] Din această perspectivă, undeva în

Attentional Modulation în the Human Visual Sistem", în A.C. Norbre și S. Kastner (coord.), The Oxford Handbook of Attention (Oxford, UK: Oxford University Press, 2015), 573 – 600; S. Treue, „Obiectând Feature-Based Attention: Monkey Physiology", în A.C. Norbre și S. Kastner (coord.), The Oxford Handbook of Attention (Oxford, UK: Oxford University Press, 2015), 601 – 619.

[0] Sunt atât de mulți oamenii de știință care au propus existența unei

procesul evoluției, sistemul nervos a devenit atât de complex, încât a depășit un prag, s-a trezit și a căpătat subiectivitate. Dacă așa ar sta lucrurile, atunci provocarea ar fi descoperirea pragului – o idee întotdeauna discutabilă.

Putem porni de la presupunerea că peștii sunt niște roboți lipsiți de inteligență care nu pot avea conștiință. Apoi un ihtiolog ne corectează prezentându-ne comportamentul lor neașteptat de bogat[0]; ajungem astfel la concluzia că, la urma urmei, și peștii au conștiință. Albinele nu sunt decât niște insecte cu un creier minuscul, dar și ele au o asemenea complexitate operațională, încât, sub anumite aspecte, rivalizează cu câinii și pisicile.[0] Cunosc pe cineva care ține o tarantulă ca animal de companie și insistă că acest animal are un comportament atât de complex, are o personalitate și un temperament atât de accentuate, încât trebuie neapărat să aibă și conștiință. Apoi în discuție încep să se amestece o mulțime de înclinații afective tipic omenești. Dacă gândacii au conștiință, poate că n-ar trebui să-i strivim, așadar e mai convenabil să tragem linie un pic mai sus pe scara complexității nervoase și să presupunem că de fapt gândacii nu au conștiință – rămânem, în final, cu niște contradicții datorate iraționalității specifice nouă, oamenilor.

Problema cu argumentul complexității este că, cu cât îl cercetezi mai îndeaproape, cu atât devine mai alunecos și mai arbitrar, mai supus capriciului omenesc. Animalele unicelulare au propriile metode complexe de a procesa informațiile, prin semnale chimice interne care interacționează unele cu altele. Dacă ești specialist în amibe și îți petreci viața scrutându-le la microscop, descoperind amănuntele traiului lor, s-ar putea să devii

---

legături între complexitate și conștiință, încât a devenit un clișeu științifico-fantastic. Cu toate acestea, Giulio Tononi a propus forma cea mai sistematică, matematică, a ipotezei. G. Tononi, Phi: A Voyage From The Brain To The Soul (New York: Pantheon, 2012).

[0] R. Bshary, W. Wickler și H. Fricke, „Fish Cognition: A Primate's Eye View", Animal Cognition 5 (2002): 1-13.

[0] C. Koch, „Exploring Consciousness Through the Study of Bees", Scientific American (2008).

campion în lupta pentru demonstrarea conștiinței amibelor. Plantele au propriile sisteme electrochimice de transmitere a semnalelor, procesând informații despre lumea din afară, așa că nu e exclus să fie conștiente. Genomul din nucleul unei celule, alături de masiva mașinărie chimică responsabilă de codul genetic, este un computer care procesează informații destul de complicate. Poate că nucleul celulei are propria conștiință.

Dar de ce să ne oprim aici? Pietrele, și apa, și electronii mustesc de informații și își schimbă starea în funcție de mediu. De ce nu sunt și acestea conștiente?

Odată ce mergi pe ideea potrivit căreia conștiința izvorăște în mod natural din procesarea complexă a informației, e dificil să nu aluneci în panpsihism, convingerea că tot ce se află în univers este conștient într-o măsură mai mică sau mai mare.[0] Dacă punem semnul egal între conștiință, informație și complexitate, proprietăți care sunt efectiv peste tot și în orice, nu ne rămâne decât să alunecăm pe această pantă.

Eu nu spun că cortexul cerebral susține conștiința pentru că este cea mai complexă mașinărie din universul cunoscut – deși este. Motivul este mult mai specific.

Să zicem că te uiți la un obiect banal – un măr. E pe o masă din fața ta și îl vezi clar. Informația vizuală despre măr trece prin sistemul tău de procesare, de la retină la cortex. Să presupunem că semnalele referitoare la măr câștigă rundele eliminatorii ale competiției. Dacă analizăm îndeaproape ierarhia ta corticală, descoperim că informațiile despre măr sunt procesate în profunzime, în timp ce semnalele concurente sunt eliminate. Informațiile despre măr ajung la rețelele cerebrale centrale, la rețelele parietal-frontale. Zeci de mii de alte pachete de informații au pierdut competiția pentru moment. Sunetele și imaginile din jurul tău, senzația hainelor pe corpul tău, amintirile, ideile, emoțiile, toate au pierdut în fața mărului, în acest moment din timp, imaginea vizuală a mărului, și poate a câtorva alte obiecte, domină mașinăria corticală a atenției

[0] D. Skrbina, Panpsychism în the West (Boston, MA: MIT Press, 2005).

tale, în timp ce alte informații concurente sunt procesate într-o măsură mult mai mică, în regiuni mult mai limitate ale cortexului tău.

Conform uneia dintre principalele teorii ale conștiinței, numite teoria spațiului global de lucru (*global workspace theory*)[0], asta-i tot. Avem o explicație completă a experienței conștiente a mărului. Voi discuta din nou această teorie într-un alt capitol, când voi compara teoria schemei atenției cu unele dintre principalele alternative și voi susține faptul că multe dintre aceste teorii nu sunt rivale, ci sunt mai degrabă interconectate, în capitolul de față vreau să vorbesc despre teoria spațiului global de lucru pentru că oferă o perspectivă utilă asupra atenției la nivelul cortexului cerebral.

În cadrul teoriei spațiului global de lucru, informația trece prin sistemul cortical. Este selectată și amplificată de către atenție până când depășește un prag și influențează o vastă rețea cerebrală. A câștigat lupta finală din competiția pentru atenție care se desfășoară în creier. Faptul că ajunge la această stare de influență masivă înseamnă că a intrat în spațiul global de lucru. Conform acestei teorii, informația din spațiul global de lucru *este* informația din conștiință. Motivul pentru care spațiului global de lucru îi este asociată proprietatea conștiinței nu este explicat. Această teorie este descriptivă, dar nu explicativă.

Pentru a demonstra că teoria spațiului global de lucru este incompletă, să vorbim cu creierul tău, ca în exemplul de mai sus, despre acest moment specific. În capitolele anterioare, în cazul animalelor nonverbale m-am bazat pe Vorbinator 5.000, mașinăria ipotetică la care se poate conecta un creier pentru a traduce în limbaj informațiile pe care le deține. Aici nu avem nevoie de acest dispozitiv ipotetic, dar doresc să mă asigur că respectăm aceeași logică. Mașinăria vorbirii din creierul uman poate verbaliza informația conținută în spațiul global de lucru. Faptul că

---

[0] B.J. Baars, A Cognitive Theory of Consciousness (New York: Cambridge University Press, 1988); S. Dehaene, Consciousness and the Brain (New York: Viking Press, 2014).

vorbirea depinde de informația internă poate părea evident, dar e ușor să fie înțeles greșit. Intuiția noastră obișnuită ne spune că atunci când oamenii vorbesc nu fac altceva decât să-și exprime experiențele interioare. Dar, în realitate, vorbirea este un tip de output de informație. Dacă o anumită informație lipsește din sistemul interior, aceasta nu poate fi exprimată verbal.

Așadar, dacă te întreb:

— Ce e în fața ta?

Poți să răspunzi:

— Un măr.

Dacă insist să-mi dai detalii, mi le poți oferi – culoare, textură, formă, poziție. Toate aceste informații sunt prezente în spațiul tău global de lucru.

Dar să presupunem că te întreb:

— Ești conștient de măr? Ai o experiență subiectivă?

Aici dăm de bucluc. Până acum, când am vorbit despre atenția corticală, nu am explicat modul sau motivul pentru care spațiul tău global de lucru, sau orice parte a cortexului, ar conține informații despre conștiință. Conține informații despre măr, așadar poate oferi detalii vizuale; dar pe ce bază ai putea oferi un răspuns la întrebarea despre experiența conștientă? însăși noțiunea de conștiință ar fi lipsită de sens pentru tine.

Ne lipsește încă un element esențial pentru a explica conștiința. După cum voi arăta în continuare, trebuie să adăugăm o schemă a atenției.

Am explicat modul în care cortexul este, în esență, o mașinărie a atenției. Dar nu are niciun sens să ai atenție dacă aceasta umblă aiurea sau este direcționată exclusiv de elemente din exterior, ca intensitatea luminii sau a sunetelor. Are nevoie de un sistem de control intern, iar un sistem de control nu poate opera fără un model intern. Schema atenției este acest model intern – o serie de informații despre procesul atenției în sine.

Pentru a explica importanța schemei atenționale, voi folosi o analogie care s-ar putea să pară la început puțin

forțată, dar care se va dovedi întemeiată. Mă gândesc la o scenă memorabilă din filmul *Butch Cassidy and the Sundance Kid*. The Kid, interpretat de Robert Redford, dă o probă pentru a deveni ajutor de șerif. I se cere să tragă la țintă stând drept, fără să se miște, postură considerată potrivită. În mod ciudat, nu nimerește. Țintește groaznic. În sfârșit, frustrat, spune: „Pot să mă mișc?” Apoi, clătinându-se din tot corpul, cu arma legănându-se în aer, trage două gloanțe și amândouă nimeresc ținta. „Mă descurc mai bine dacă mă mișc”, mormăie el.

Această scenă este un bun exemplu de tehnică a controlului. E preferabil să fii cât mai liber și mai mobil cu putință. Nu vrei să monitorizezi și să controlezi totul. Nu ai nevoie de un model pentru orice detaliu. În schimb, un bun sistem de control monitorizează, influențează și determină rezultatul esențial – în cazul lui Sundance Kid, atingerea țintei.

Un alt exemplu clasic de tehnică a controlului este modul în care se bat cuiele. Cu puțin exercițiu, mai toți ne pricepem să dăm cu ciocanul și, de cele mai multe ori, nimerim cuiul. Dar cum anume controlăm ciocanul? O abordare poate fi monitorizarea și controlarea fiecărei mișcări. Umerii se rotesc, cotul se îndreaptă, încheietura se mișcă ușor în jos. Dacă optimizezi fiecare dintre aceste mișcări și o perfecționezi, vei ajunge poate un expert în lovirea cu ciocanul. Se pare însă că nu asta fac oamenii. Ce facem de fapt este să ne concentrăm asupra loviturii pe care capul ciocanului o aplică cuiului.[0]37 Monitorizăm și controlăm acest proces pentru că este singurul rezultat care contează în această sarcină. Dacă ai filma oameni care dau cu ciocanul și le-ai urmări în detaliu mișcările, ai observa că umerii se rotesc în moduri variate, cotul nu se mișcă într-un singur fel și încheietura are un unghi ușor diferit de fiecare dată, și totuși capul ciocanului reușește să lovească fără probleme cuiul (cel puțin în cazul celor care se pricep la bătut cuie). Sistemul de control nu trebuie să

[0] E. Todorov și M.I. Jordan, „Optimal Feedback Control as a Theory of Motor Coordination”, Nature Neuroscience 5 (2002): 1226 – 1235.

influențeze sau să determine detaliile atâta timp cât poate controla rezultatul esențial – lovirea directă a cuiului, în acest caz. Detaliile pot să varieze oricât cu condiția ca toate laolaltă să ducă la rezultatul potrivit.

Modul în care Sundance Kid trage la țintă și arta subtilă de a bate un cui sunt ambele exemple ale modului în care creierul folosește modele interne. Creierul trebuie să creeze modele pentru ceea ce e esențial, nu pentru detaliile întâmplătoare. Altfel, așa cum Sundance Kid trebuia să fie atent la prea multe detalii intermediare, sistemul de control s-ar prăbuși și n-ar mai putea să-și îndeplinească sarcina. Când creierul controlează atenția, menținând-o într-un punct sau schimbându-i direcția, mai rapid sau mai lent, lărgindu-i spectrul de interes sau dimpotrivă, ce fel de model intern utilizează? Ce aspecte ale atenției sunt simulate de acest model?

În cazul atenției fățișe, pe care am descris-o în capitolul anterior, creierul poate ține socoteala atenției monitorizând un organ fizic – ochiul, de exemplu, în timp ce se mișcă și se îndreaptă spre diverse obiecte. Dar în cazul atenției mascate nu este implicat niciun organ fizic. Când atenția mascată trece de la un element la altul – de la măr la un sunet, apoi la o amintire –, nu se mișcă nimic tangibil în spațiu. Însă miliarde de neuroni își schimbă starea de activare în moduri subtile. La nivel microscopic, atenția fățișă rămâne un proces fizic, dar o schemă atențională utilă nu ar modela detaliile neuronilor, conexiunile, inhibitorii, rundele eliminatorii ale competiției, ierarhiile corticale sau rețelele lobilor parietal și frontal. Creierul nu are nevoie să cunoască toate detaliile interne, mecanice, pe care le-am subliniat până aici în acest capitol. În schimb, pentru a fi utilă, schema atențională trebuie să reprezinte ceva simplificat care se reduce la esența pragmatică a atenției mascate.

Ajungem acum la afirmația centrală a teoriei. Colegii mei și cu mine propunem ipoteza că schema corticală a atenției are o anumită formă. Informația pe care o conține oferă o ilustrare simplificată a modului în care nivelurile

superioare ale atenției corticale intră în posesia elementelor din câmpul vizual. Nu are loc aici o mișcare simplă, fizică, a ochiului, cum se întâmplă în cazul atenției fățișe. În schimb, ilustrarea simplificată descrie o esență care nu are o anumită substanță fizică, dar are o poziție vagă înăuntrul tău care poate intra temporar în posesia unor elemente – mere, sunete, gânduri și amintiri – și care se mișcă neîncetat, căutând, asimilând unele informații și renunțând la altele. Când această esență mintală eterică ia în posesie un element, ea are proprietatea de a-l face clar pentru tine, real, conferindu-i o prezență intensă – cu alte cuvinte, transformă acel element într-o experiență. Iar asta are consecințe importante. Îți permite să înțelegi acel element, să reacționezi la el, să vorbești despre el, să ți-l amintești astfel încât să poți lua hotărâri mai târziu. Îți dă *puterea* să acționezi.

Această putere amorfă dinăuntrul tău este o exemplificare fictivă, săracă în detalii, a atenției corticale. Nu mă refer la atenția de nivel scăzut, cum ar fi competiția dintre semnalele vizuale din V1, ci la atenția de la cel mai înalt nivel, în care un lucru cum ar fi un măr poate câștiga finala campionatului și poate avea un efect asupra comportamentului tău.

Când te întreb: „Care este relația ta mintală cu acel măr?”, mașinăria ta verbală poate accesa informația disponibilă în rețelele tale corticale, informații care au ajuns în spațiul tău global de lucru. Nu se bazează doar pe informații vizuale despre măr, ci și pe informații din schema atențională despre o putere amorfă dinăuntrul tău. Aceste două pachete de informații sunt interconectate și formează un întreg, un fel de dosar care conține toate informațiile relevante în acel moment cu privire la măr. Pe baza acestor informații poți spune: „Când mă uit la acest măr, îl dețin la nivel mintal, am *experiența conștientă* a culorii sale roșii”.

Să presupunem că te întreb: „Dar la ce te referi prin experiență conștientă? Care sunt proprietățile ei specifice, fizice?” Nu poți răspunde cu ușurință la această întrebare. Schema atențională, fiind săracă în detalii, nu are o

descriere a vreunei proprietăți fizice specifice a atenției.

Eu întreb:

— Poți să atingi experiența conștientă și să vezi cât e de tare? Poți să o pui pe cântar și să vezi ce greutate are? Poți să o încălzești și să vezi care e temperatura ei de combustie? Ce măsurători fizice i se pot aplica?

Ai putea răspunde:

— Nu are niciuna dintre aceste caracteristici fizice. În acest sens, este non-fizică sau metafizică. Este pur și simplu o experiență mintală – este modul în care mintea mea înțelege ceva. Nu știi ce e conștiința?

Logic, creierul nu poate să spună ceva decât dacă conține informații pe care să se bazeze. În centrul teoriei schemei atenționate este pachetul de informații pe care se bazează afirmația despre experiența subiectivă. Dat fiind că creierul nostru construiește un model schematic al atenției corticate, știi ce este conștiința și crezi că o ai. Poți să răspunzi la întrebări când îți sunt adresate și, când citești o carte despre conștiință, cum ar fi aceasta, știi într-o anumită măsură la ce mă refer. Fără o schemă atențională ți-ar lipsi informația necesară pentru a face oricare dintre aceste lucruri, iar conștiința nu ar avea sens sau ar fi irelevantă pentru tine.

Conform perspectivei mele, teoria schemei atenționate a conștiinței are o logică inevitabilă. În primul rând, știm că cortexul folosește atenția fățișă. În al doilea rând, știm că trebuie să controleze această atenție. În al treilea rând, știm că creierul trebuie să aibă un model intern al atenției pentru a o putea controla. În al patrulea rând, știm că un model intern detaliat, foarte exact, ar fi în cel mai bun caz o risipă de resurse și ar influența negativ procesul, așa că acest model intern al atenției trebuie să nu ia în considerare detaliile tehnice. Așadar, și în al cincilea rând, o schemă atențională ar reprezenta sinele ca având o putere internă amorfă, non-fizică, capacitatea de-a ști, de-a experimenta și de-a reacționa, de-a se concentra asupra unor elemente disparate – esența atenției mascate, fără detaliile care stau la baza ei. Pornind de la principiile prime,

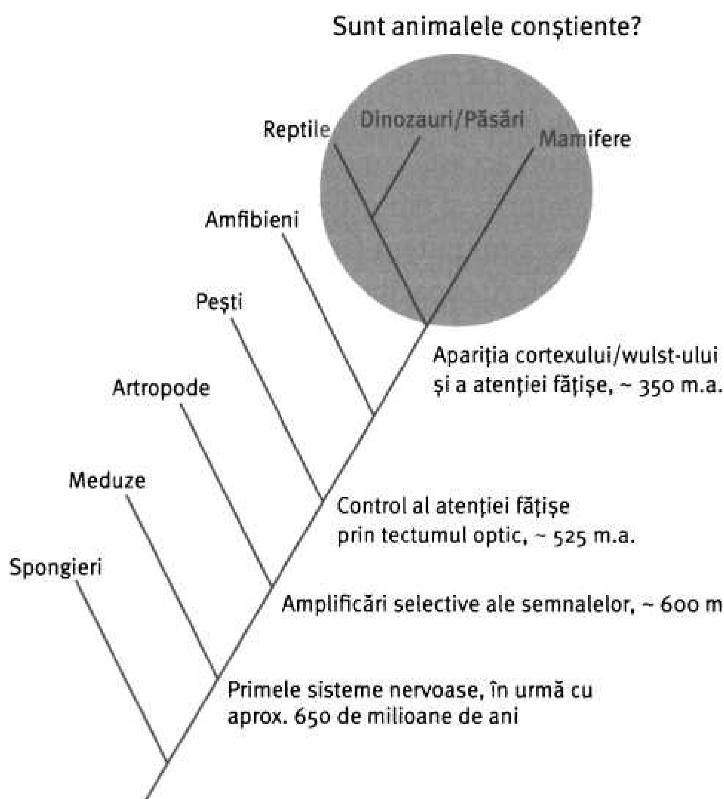
dacă ar trebui să construiești un creier care să funcționeze bine, cu o puternică atenție mascată, ai construi o mașină care, bazându-se pe informațiile disponibile în interior, ar afirma că are o conștiință non-fizică.

Bineînțeles, această mașinărie corticală nu ar ști că propria experiență a conștiinței subiective este un construct sau o simplificare. Ar considera reală natura non-fizică a experienței conștiente, pentru că – oarecum tautologic – creierul nu știe decât ceea ce știe. Este captiv propriilor informații.

Să ne gândim la toate aspectele cerebrale despre care am vorbit în ultimele trei capitole și care nu sunt de ajuns pentru explicarea conștiinței. Complexitatea singură nu este de ajuns. Cel mai complex creier din lume s-ar putea să nu aibă nicio informație relevantă pentru propria conștiință. Nu este suficient nici tipul fățiș de orientare a atenției, cel controlat de tectumul broaștei. O broască își poate îndrepta capul spre o muscă sau în direcția unui pericol, fără să aibă vreo informație relevantă pentru conștiință. Nici măcar atenția sofisticată, corticală, nu este de ajuns. Când informația referitoare la măr câștigă competiția corticală și ajunge în spațiul global de lucru, această stare nu spune nimic despre conștiință. Nu există niciun motiv logic pentru care o mașinărie de acest tip să știe ceva despre conștiință, să se gândească că are conștiință sau să facă afirmații despre conștiință. Mai e nevoie de ceva. E nevoie de un model intern care să reprezinte atenția corticală. Această schemă atențională pare o adăugire minoră, fără de care însă sistemul nu are informația necesară pentru a revendica experiența subiectivă. Odată ce este prezentă, conștiința devine în sfârșit relevantă pentru mașinărie.

Toată lumea vrea să știe ce animale sunt conștiente. În ultimele capitole am prezentat unele componente ale conștiinței care ar fi putut exista încă din trecutul îndepărtat, începând cu o jumătate de miliard de ani în urmă. Figura 4.2 ilustrează modul în care aceste

componente s-au format, una câte una, de-a lungul evoluției. Dacă teoria schemei atenționale este corectă, atunci conștiința așa cum o înțelegem noi, oamenii, probabil că a apărut demult, în urmă cu aproximativ 300 de milioane de ani. Structurile cerebrale elementare au apărut prima oară la reptile, sunt probabil prezente la păsări și sunt cu siguranță prezente la mamifere. Deși poate varia gradul de sofisticare sau bogăția atenției fățișe și a schemelor atenționale la aceste trei grupuri, e posibil totuși ca toate să aibă o versiune a ceea ce noi recunoaștem drept conștiință.



4.2 Evoluția probabilă a conștiinței, de la spongieri la mamifere.

Însă nu am încheiat povestea evoluției. Nu am ajuns încă la partea cea mai umană. Noi nu doar construim

modele descriptive bogate ale propriei persoane, ci ne și atribuim automat conștiință unii altora, creând astfel un ecosistem social, înzestrăm cu conștiință oamenii, animalele de companie, jucăriile, vasta lume invizibilă a zeilor și spiritelor pe care le proiectăm în spațiul din jurul nostru. În capitolul următor, voi aborda modul esențial în care utilizăm social conștiința, atât de radical amplificat în cazul speciei noastre.

5
Conștiința socială

Noi, oamenii, nu ne dăm pur și simplu cu presupusul în privința stărilor psihice ale celorlalți. Nici nu facem analize complexe ca să ne citim unii pe alții. Poate uneori, dar rezultatul nu e grozav. Ce facem de fapt este să ne folosim intuiția îndelung șlefuită. Avem impresia că *știm* la ce se gândesc și ce simt ceilalți. Uneori această cunoaștere este atât de palpabilă, încât simțim gândurile și emoțiile celorlalți ca pe niște emanații. Bineînțeles că nu putem face asta. Dar milioane de ani de evoluție ne-au oferit capacitatea de a descifra indicii subtile și de a construi modele bogate ale minții celorlalți – ceea ce facem mai degrabă intuitiv decât explicit.

Ne atribuim unii altora emoții, intenții, planuri, convingeri – întregul spectru al conținuturilor mintale. Acest proces uman minunat de complex prin care reconstruim mintea altcuiva se numește *teorie a minții*.[0] Nu este o teorie în sens intelectual; este automată și inevitabilă. Nu ne putem abține. Însă, când construim aceste modele ale minții celorlalți, un anumit element are o importanță

---

[0] MJ. Doherty, How Children Understand Others’ Thoughts and Feelings (New York: Psychology Press, 2008); U. Frith și C.D. Frith, „Development and Neurophysiology of Mentalizing”, Philosophical Transactions of the Royal Society of London, B Biological Sciences 358 (2003): 459 – 473; D. Premack și G. Woodruff, „Does the Chimpanzee Have a Theory of Mind?”, Behavioral and Brain Sciences 1 (1978): 515 – 526.

aparte: reconstruirea atenției celuilalt.

Cum aș putea să știu dacă ai de gând să apuci acel măr fără să știu mai întâi dacă ești atent la el? Și, chiar dacă știu că o să iei mărul, cum aș putea să fac predicții despre ce o să faci sau o să spui apoi dacă nu înțeleg consecințele atenției? Primul pas pe care îl fac când reconstruiesc mintea cuiva este să înțeleg că o minte este înzestrată cu atenție, că aceasta poate avea spectru larg sau restrâns, în funcție de circumstanțe, că poate trece de la un obiect la altul, ceea ce are consecințe previzibile. Dacă nu aș ști toate astea, nu aș avea o teorie a minții. Am nevoie de mai mult decât de o teorie a *conținutului* minții tale; am nevoie în primul rând de o teorie despre ce *este* mintea.

Dacă stau în fața ta, îmi e simplu să văd în ce direcție privești. O întreagă știință a fost construită în jurul modului în care creierul procesează direcția privirii altcuiva.[0] Dar simplul fapt de a urmări direcția atenției tale fățișe nu-mi este deloc suficient pentru a-ți reconstrui starea mintală. Trebuie să înțeleg atenția ta *ascunsă*. Trebuie să mă folosesc de orice indiciu la îndemână și să țin seama de întreg contextul, inclusiv de limbajul tău corporal, de expresiile tale faciale, de cuvintele tale și de cunoașterea mea generală despre tine. Indiferent în ce direcție se întâmplă să privești și fie că-ți pot vedea ochii sau nu, trebuie să reconstruiesc elementele care se poate să fi pătruns în ierarhiile tale corticale, ajungând la cele mai înalte niveluri de procesare, și trebuie să înțeleg modul în care aceste niveluri superioare de procesare îți pot influența comportamentul.

Pot să garantez că nimeni nu s-a uitat vreodată la

---

[0] S. Baron-Cohen, Mindblindness: An Essay on Autism and Theory of Mind (Cambridge, MA: MIT Press, 1997); C.K. Friesen și A. Kingstone, „The Eyes Have it! Reflexive Orienting is Triggered by Nonpredictive Gaze”, Psychonomic Bulletin and Review 5 (1998): 490 – 495; E.A. Hoffman și J.V. Haxby, „Distinct Representations of Eye Gaze and Identity în the Distributed Human Neural Sistem for Face Perception”, Nature Neuroscience 3 (2000): 80 – 84; L.A. Symons, K. Lee, C.C. Cedrone și M. Nishimura, „What Are You Looking At? Acuity for Triadic Eye Gaze”, Journal of General Psychology 131 (2004): 451 – 469.

altcineva și a fost fulgerat de gândul: „Căile vizuale corticale ale acestei persoane sunt angajate în procesarea a numeroși stimuli în acest moment, neuronii care reprezintă forma mărului având o activitate mai intensă datorită amplificării semnalelor determinate de regiuni ale lobilor frontali, consecința fiind că acești neuroni au inhibat parțial activitatea neuronilor din apropiere prin inhibare laterală dependentă de interneuronii locali care utilizează acidul gama-aminobutiric...” Aș putea continua cu acest jargon. Dar adevărul e că nimeni nu a atribuit vreodată altcuiva atenție concretă, fizică, neuronală. Nu avem nevoie să facem asta și nu e deloc necesar să ajungem la asemenea detalii. Creierul meu construiește însă un model mult mai schematic și mai eficient. Înțeleg în mod intuitiv că: „În acest moment conștiința acestei persoane pare să se concentreze asupra mărului, cu numeroase consecințe posibile”.

Ajungem iar la conștiință ca model simplificat și util al atenției. Dar, de data asta, schema atenției se află în serviciul inteligenței sociale: pentru a construi un model al altcuiva, nu al propriei persoane.

Sally și Anne se plimbă prin parc și fiecare are un coș de picnic acoperit cu o pânză. Sally pune un sandviș în coșul A și se duce la toaletă. Cât timp e plecată, Anne pune pe furiș sandvișul în coșul B și acoperă din nou primul coș. Când Sally se întoarce, în ce coș se va uita mai întâi pentru a găsi sandvișul? Această întrebare simplă a devenit primul test pentru verificarea capacității de construire a unei teorii a minții.[0]

Pentru a trece testul trebuie să urmărești informațiile din mintea lui Sally. E corect ce știe despre locul în care se află sandvișul când îl pune în coșul A, dar apoi această

[0] S. Baron-Cohen, A.M. Leslie și U. Frith, „Does the autistic child have a «theory of mind?»”, Cognition 21 (1985): 37 – 46; H. Wimmer și J. Ferner, „Beliefs About Beliefs: Representation and Construining Function of Wrong Beliefs în Young Children’s Understanding of Deception”, Cognition 13 (1983): 103 – 128.

informație devine incorectă când sandvișul e mutat în coșul B. Nu poți să rezolvi această sarcină dacă nu înțelegi că mintea lui Sally e o entitate separată ce conține informații, posibil false, care îi determină acțiunile. Din acest motiv testul este numit uneori sarcina convingerilor false. Răspunsul corect este: se va uita mai întâi în coșul A și va descoperi că sandvișul lipsește.

Copiii mai mici de cinci ani nu pot să ofere în mod consecvent răspunsul corect.[0] Ei cred că, dacă sandvișul e în coșul B, Sally se va uita acolo. De ce ar deschide un coș în care nu se află ce caută ea? Pe măsură ce cresc, gândirea lor socială se ascute și sarcina devine intuitivă și simplă. Când ajungem la vârsta adultă, ne pricepem în general destul de bine să ținem seama de stările mintale ale celorlalți.

Cimpanzeii prezintă unele dovezi că ar putea rezolva sarcina convingerilor false.[0] Să presupunem că derulezi scenariul Sally-Anne în fața unui cimpanzeu. Sub privirea cimpanzeului, Sally pune un fruct într-o cutie și pleacă. Anne mută fructul în cutia B. Apoi Sally se întoarce și caută fructul. Dacă te uiți la direcția în care privește cimpanzeul, observi că tinde să se uite mai degrabă la cutia A, unde Sally crede că e fructul și unde ar trebui să caute mai întâi, și nu la cutia B, unde fructul se află de fapt. Probabil că cimpanzeul urmărește conținutul minții lui Sally și îi anticipează acțiunile.

E posibil ca și corbii să poată rezolva o sarcină similară.[0] Aceste păsări obișnuiesc să pună deoparte hrană pentru mai târziu și nu le place când alte păsări le-o fură. Să presupunem că o pasăre ascunde o bucățică de hrană în

---

[0] H.M. Wellman, D. Cross și J. Watson, „Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief”, Child Development 72 (2001): 655 – 684.

[0] C. Krupenye, F. Kano, S. Hirata, J. Call și M. Tomasello, „Great Apes Anticipate That Other Individuals Will Act According to False Beliefs”, Science 354 (2016): 110 – 114.

[0] N.S. Clayton, „Ways of Thinking: From Crows to Children and Back Again”, Quarterly Journal of Experimental Psychology 68 (2015): 209 – 241.

timp ce o altă pasăre o privește. Apoi pasărea care privește zboară. Cât timp aceasta lipsește, pasărea care ascunde hrana o mută cu șiretenie în altă parte, ca să nu-i fie furată. Se pare că această pasăre înțelege că o alta a aflat unde e ascunsă hrana și că atunci când se va întoarce va avea o convingere falsă și o va căuta într-un loc în care hrana nu se mai află.

Nu multe animale nonumane pot trece testul convingerilor false. Chiar și cele câteva excepții pe care le-am prezentat sunt controversate. Dar, în opinia mea, este prematură concluzia că alte animale nu au o teorie a minții. Sarcina convingerilor false este un standard prea înalt. Să urmărești mai multe cutii și schimbările de conținut din ele este prea complicat sub aspect intelectual, ca un joc dificil de alba-neagra. Nu e de mirare că doar oamenii pot rezolva consecvent această sarcină. Eu sunt interesat de ceva mai simplu: conceptul de minte. Cu toții înțelegem că Sally are o minte și că mintea este ceva ce conține informații și poate determina comportamentul pe baza acelor informații. Au alte animale aceeași înțelegere intuitivă? Știu ele ce înseamnă pentru altcineva să fie conștient de ceva?

Oamenii de știință care studiază comportamentul animalelor preferă explicații simple. În loc să presupunem că un animal are vreo concepție despre conștiința altcuiva, pare mai simplă presupunerea că animalul nu a făcut decât să învețe o serie de reguli simple. De exemplu, poate că o zebră nu are nevoie să știe că un leu a văzut-o. În schimb, zebra poate pur și simplu să fugă de orice animal mare înzestrat cu colți. Un simplu răspuns la un stimul. Dacă zebra are o bază de date suficient de mare cu astfel de asocieri, se descurcă probabil. Eu cred însă că o astfel de ipoteză, specifică psihologiei de tip stimul-răspuns, este de fapt naivă. O bază de date mare, cu multe asocieri, nu este metoda cea mai ușoară sau cea mai eficientă de-a te mișca printr-un mediu complicat. Din perspectivă computațională, o abordare bazată pe modele este probabil mai simplă, dat fiind că un model poate servi unei game largi de circumstanțe. Ar putea fi mai simplu și mai economic

pentru zebră să construiască un model schematic conform căruia leul are o minte, aceasta poate absorbi informații din lume și îi poate determina astfel acțiunile.

Poate părea neîntemeiată presupunerea că zebra „înțelege” conștiința altui animal, dar asta se întâmplă doar pentru că noi, oamenii, considerăm conștiința o caracteristică nobilă, specifică nouă și bogată în asociații culturale. Zebrele nu pot avea capacitatea noastră pentru poezie și complexitate. Așa ne spune vanitatea noastră. Eu cred că conștiința este o parte străveche a teoriei minții; este un model simplu și eficient, utilizat pentru a face predicții referitoare la comportamentul animalelor, și e probabil să fi evoluat cu mult înainte de apariția oamenilor. N-aș fi surprins dacă zebrele, alte mamifere, păsările și poate chiar unele reptile ar folosi constructe convenabile ale conștiinței, cu niveluri variate de complexitate, pentru a-și prezice unele altora comportamentul.

E un clișeu că noi, oamenii, avem o conștiință superioară. Avem tendința să credem că alte animale nu au deloc conștiință sau că au o conștiință mai puțin dezvoltată. Această perspectivă se potrivește de minune cu ideea larg răspândită conform căreia conștiința apare din complexitate. Pentru că avem un creier mai complex decât al celorlalte animale înseamnă că avem și o conștiință mai complicată. Însă dintre toate abilitățile mintale cu care ne place să ne lăudăm ca specie – matematica, limbajul, folosirea uneltelor și așa mai departe – conștiința ar putea fi una dintre cele mai primitive și mai puțin speciale. Nu neg faptul că conținutul conștiinței, gândurile, și ideile, și credințele, intuiția profundă, înțelegerea mortalității sunt probabil mai sofisticate în cazul oamenilor decât al celorlalte animale. Dar conștiința însăși, capacitatea noastră de a avea o experiență subiectivă și de-a atribui o experiență subiectivă altora are o utilitate atât de bazală, încât poate fi comună foarte multor specii. Dacă teoria schemei atenționale este corectă, atunci cu siguranță nu se aplică doar oamenilor.

Pentru o înțelegere mai profundă a lucrurilor, găsesc adesea utilă perspectiva tehnică. Să ne gândim la cum se construiește o mașinărie care poate prezice comportamentul uman de la un moment la altul. Acest exercițiu ne va arăta cât e de dificilă sarcina pe care o rezolvă creierul și în același timp va dezvălui unele principii fundamentale. Ne va indica și calea spre capitolele de la sfârșitul cărții, unde voi discuta cât este de fezabil să construim vreodată conștiință artificială.

Să presupunem că un bărbat (să-i spunem Kevin) intră într-o încăpere. O cameră îl urmărește și un microfon îl ascultă fără ca el să știe. Camera și microfonul sunt conectate la un computer, Predictor 5.000, a cărui sarcină este să ofere predicții despre ce va face Kevin.

În încăpere putem vedea: o gogoașă pudrată cu zahăr, în mijlocul unei mese aflate în centrul încăperii, cu o lumină care strălucește deasupra gogoșii; apă vărsată pe jos, în dreptul mesei; și un telefon mobil pe un raft din colțul încăperii, unde lumina e slabă.

Prima sarcină pe care i-o vom da mașinăriei este să identifice disponibilitățile (*affordances*) încăperii. Influentul psiholog și naturalist J. J. Gibson a introdus termenul *affordance* în anii 1970. El a înțeles că, atunci când un animal sau un om se uită în jur, adevărata sarcină a sistemului vizual nu este să perceapă lumea așa cum e, ca și cum ar face o poză, ci să identifice oportunitățile de acțiune. Aceste oportunități sunt numite *affordances*, O muscă reprezintă o disponibilitate pentru o broască, întrucât broasca poate să o mănânce. O creangă reprezintă o disponibilitate pentru o pasăre, întrucât pasărea se poate așeza pe ea. O clanță reprezintă o disponibilitate pentru un om, întrucât o poate apăsa.

Devine imediat limpede cât de complex trebuie să fie un mecanism funcțional de predicție a comportamentului. Predictorul 5.000 trebuie să se priceapă la comportamentul uman pentru a putea să vadă persoana, să vadă apa și apoi să actualizeze informația că persoana poate ocoli balta. Să ne imaginăm această sarcină repetată pentru fiecare obiect

peste care ar putea da cineva. Un mecanism funcțional de predicție ar trebui să cunoască un volum enorm de detalii despre comportamentul uman. Deși nu este neapărat complicată punerea în aplicare a acestui tip de cunoaștere în inteligența artificială – e o chestiune de recunoaștere a obiectelor și de asociere a acestora cu acțiuni potențiale –, există totuși un număr uriaș de asociații care trebuie învățate.

Pentru a face această problemă și mai dificilă, fiecare obiect poate avea un număr mare de disponibilități. Încăperea poate părea spartană, nefiind acolo decât trei obiecte, dar are multe posibilități ascunse. Kevin poate păși peste baltă; poate să sară în ea și să stropească în jur; sau poate să scoată o batistă și să o șteargă. Poate să răstoarne masa într-o criză de furie; poate să o mute atent într-o altă parte a încăperii; poate să se urce pe ea; sau poate să se ascundă sub ea. Poate să ia gogoașa și să o mănânce; poate să o arunce pe jos și să o calce în picioare; sau, în glumă, poate să și-o pună la ochi și să se uite prin ea ca printr-un monoclu. Poate să ia telefonul și să apeleze pe cineva; poate să și-l bage pe furiș în buzunar; sau poate să facă ceva simplu, cum ar fi să se îndrepte spre el ca să-l vadă de aproape. Mecanismul de predicție are de-a face cu un număr uriaș de posibile disponibilități.

Pentru a simplifica sarcina, îi vom oferi Predictorului 5.000 unele informații probabilistice – informații generale despre modul în care se comportă o persoană obișnuită. Majoritatea oamenilor nu sar în mod normal într-o baltă, nu răstoarnă masa și nu calcă în picioare o gogoașă. Acestea sunt evenimente puțin probabile. Cu suficiente informații despre comportamentul uman putem estima probabilitatea ca un om obișnuit să pună în practică aceste acțiuni atunci când îi iese în cale o baltă sau o gogoașă.

Dar, chiar dacă preîncărcăm mașinăria cu aceste informații, tot nu avem un bun mecanism de predicție. Informațiile de până acum se referă la un om obișnuit. Nu știm cum se va purta în aceste condiții specimenul nostru, Kevin. Să ajutăm mașinăria noastră de predicție oferindu-i

întregul aparat al unei teorii tradiționale a minții. De exemplu, dacă știu că Kevin nu a mâncat de zece ore, pot să estimez că sunt șanse mari ca el să înfulece gogoașa. Dacă știu că are diabet, pot să presupun că n-o s-o mănânce. Dacă știu că lui Kevin îi e greu să-și controleze impulsurile sau că e foarte furios în acel moment, pot să presupun că sunt mari șansele ca el să calce furios în baltă. Filosoful Daniel Dennett numește acest tip de informații „atitudine intențională”.[0] Ne uităm mereu la alți oameni și ne punem automat întrebarea: „Pot să-mi dau seama care sunt motivațiile acestei persoane ca să pot prezice ce va face în continuare?”

Cercetătorii au început să construiască sisteme artificiale care încearcă să rezolve problema predicției comportamentale utilizând atitudinea intențională.[0] Capacitatea de a identifica atitudinea intențională a altcuiva implică probabil stocarea unor cunoștințe complexe despre natura umană – despre motivațiile umane tipice și în special despre emoțiile manifestate prin expresii faciale. Dar vreau să ocolesc chestiunea atitudinii intenționale și să ajung la ceva mai simplu și, cred, fundamental.

Să presupunem că am luat în calcul informațiile statistice disponibile despre comportamentul uman obișnuit, combinate cu informații specifice despre Kevin și despre atitudinea lui intențională din dimineața asta, și am făcut din toate o serie de numere. Fiecărei disponibilități a încăperii i-a fost atribuită o probabilitate. Să mănânce gogoașa: 30%. Să ocolească atent balta: 50%. Să ia telefonul pe care altcineva l-a lăsat pe raft: 3%. Am făcut

---

[0] D.C. Dennett, The Intențional Stance (Cambridge, MA: Bradford Books/ MIT Press, 1987).

[0] C.L. Baker, R. Saxe și J.B. Tenenbaum, „Action Understanding as Inverse Planning”, Cognition 113 (2009): 329 – 349; N.C. Rabinowitz, F. Perbet, F. Song, C. Zhang, S.M. Ali Eslami și M. Botvinick, „Machine Theory of Mind”, Computer Science arxiv (2017): 1802.007740; R. Saxe și S.D. Houlihan, „Formalizing Emotion Concepts Within a Bayesian Model of Theory of Mind”, Current Opinion în Psychology 17 (2017): 15 – 21.

toată munca de cercetare care în mod normal constituie substanța cogniției sociale și i-am oferit informațiile Predictorului 5.000. Se pare că i-am făcut eu treaba.

Însă, cu toate aceste informații preîncărcate, mașinăria tot nu poate prezice comportamentul lui Kevin de la un moment la altul. Are nevoie de informații despre o variabilă esențială, ascunsă, care se modifică în timp real: cum își concentrează Kevin atenția. Resursele de procesare ale lui Kevin își schimbă constant ținta. Drept urmare, probabilitățile de a face ceva cu gogoașa, sau cu balta, sau cu telefonul variază continuu.

Să luăm cazul gogoșii. Gestul de-a arunca gogoașa pe jos și de-a o călca în picioare are o probabilitate anterioară atât de scăzută, încât această acțiune este eliminată de mașinăria de predicție. Gestul de-a o folosi ca monoclu este de asemenea eliminat. Nu ne rămâne decât o disponibilitate realistă sau probabilă în privința gogoșii: s-o ia și s-o mănânce. Dar probabilitatea acestei acțiuni fluctuează în timp. În încercarea de a prezice această fluctuație voi intra puțin în matematică, utilizând un cadru de referință numit statistică bayesiană. Dacă nu te interesează matematica, poți să sari peste partea asta, dar promit că o să fac lucrurile cât mai simple cu putință.

Să presupunem că, dat fiind tot ce știm despre oameni în general și despre Kevin în particular, în cazul lui există o probabilitate generică de a mânca gogoașa de 30%. Acest număr se numește probabilitate anterioară, și asta am preîncărcat în mașinărie. Să îi spunem Aanterioară. Acum să presupunem că mașinăria noastră de predicție calculează un al doilea număr, care la rândul său poate să varieze între 0% și 100%. Acest al doilea număr reprezintă o estimare a volumului de atenție pe care Kevin o direcționează spre gogoașă. Vom numi acest număr Cp unde 1 desemnează obiectul 1, gogoașa. Mai apoi ne vom referi la C_2 și C_3, care desemnează volumul de atenție pe care Kevin îl direcționează spre baltă și spre telefonul mobil. C_l fluctuează constant în timp. În cea mai mare parte a timpului, Kevin nu este foarte atent sau nu este deloc

atent la gogoașă, deci C_I este aproape de O. Ocazional, atenția lui se poate îndrepta spre gogoașă, și atunci C: va crește puțin; sau se poate ca atenția lui să crească brusc, și atunci C_I va urca temporar la aproximativ 100%.

Cu cât atenția lui Kevin este mai concentrată asupra gogoșii – deci cu cât C_I devine mai mare –, cu atât este mai probabil să interacționeze cu gogoașa. Ci este un tip de variabilă care permite posibilitatea de acțiune. Aici apare singura ecuație din carte. Să presupunem că întrebăm Predictorul 5.000: „Care este probabilitatea, să-i spunem Pacțiune, ca în acest moment Kevin să mănânce gogoașa?”

Mașinăria estimează probabilitatea printr-o operație simplă:

$$P_{acțiune} = C_I \times A_{anterioară}$$

Asta-i tot. Dacă înmulțești probabilitatea anterioară cu atenția estimată, ai o modalitate de a prezice comportamentul lui Kevin de la un moment la altul. Atâta timp cât nu e deloc atent la gogoașă, și $C_I = 0$, deci $P_{acțiune} = 0$, mașinăria prezice că nu o s-o mănânce. Imediat ce crește brusc atenția îndreptată spre gogoașă, probabilitatea estimată de-a o mânca va crește la rândul său. Nici chiar într-un moment de atenție maximă probabilitatea de a mânca gogoașa nu depășește probabilitatea $A_{anterioară}$ care poate fi destul de mică, dat fiind că în general oamenii nu prea mănâncă gogoși la întâmplare. Când atenția pe care o acordă gogoșii scade din nou, probabilitatea de-a o mânca va scădea și se va apropia de O. Utilitatea acestui tip de operație constă în faptul că la abordarea standard a teoriei minții, care tinde să opereze ca un stop-cadru înghețat în timp – dacă lui Kevin i se oferă o gogoașă, o va mânca? —, și o pune într-un cadru de referință care ține seama de schimbările continue, dinamice, de la o secundă la alta, ale atenției.

Sarcina mașinăriei este de a estima acest număr, C_I aflat constant în schimbare. Dar mecanismul nostru artificial de predicție nu are acces direct la creierul lui Kevin. Atenția lui Kevin este rezultatul unei serii foarte complexe de interacțiuni neuronale ascunse în craniul său. Mașinăria nu

are decât o cameră care îl urmărește pe Kevin și un microfon care înregistrează sunetele din încăpere. Are nevoie de niște operații simple care pot transforma observațiile sale limitate într-o estimare a nivelului de atenție a lui Kevin.

Pentru a ajuta mașinăria noastră, să ne folosim de câteva cunoștințe demonstrate științific. În primul rând, gogoașa e albă și e luminată direct – are un potențial înalt de stimulare, tinde să iasă în evidență. Faptul că gogoașa iese în evidență înseamnă că are mai multe șanse să atragă atenția. Așadar, mașinăria poate crește valoarea estimată a lui C_l în al doilea rând, gogoașa este singurul lucru de pe masă, iar atenția depinde de numărul de obiecte sau de competiția vizuală. Din nou, pe baza acestui indiciu, mașinăria poate crește valoarea estimată a lui C_l în al treilea rând, direcția în care privește Kevin oferă informații utile. Direcția privirii nu este un indicator perfect – s-ar putea ca el să se uite la gogoașă și de fapt să fie preocupat de altceva, cum ar fi să asculte cu atenție un zgomot care se aude din holul din spatele lui sau să se gândească intens la planurile pentru mâine. Dar, în general, direcția privirii oferă informații importante atunci când se încearcă estimarea atenției cuiva. În al patrulea rând, expresia lui facială oferă și ea indicii. Fața i se schimbă brusc, trecând de la o expresie neutră la una mai concentrată imediat ce privirea îi cade asupra gogoșii, mașinăria având astfel motive să calculeze că valoarea lui C_l tocmai a crescut.

Luând în calcul toate aceste considerații – relația atenției cu acele calități ale obiectelor care ies în evidență, cu numărul de obiecte, cu direcția privirii și cu expresia facială –, mașinăria poate estima o valoare variabilă în timp a lui C_l volumul de resurse de procesare pe care Kevin le direcționează spre obiectul 1. Această estimare fluctuează în timp, pe măsură ce mașinăria asimilează informații schimbătoare. Utilizând valoarea lui Cp poate calcula probabilitatea ca Kevin să mănânce gogoașa: în acest moment s-ar putea să o mănânce; în momentul ăsta nu; cu

siguranță nu; nu; da, acum este din nou posibil.

Situația devine mult mai interesantă dacă luăm în considerare alte obiecte din încăpere. De exemplu, apa de pe jos. Mașinăria calculează o valoare pentru C_2, care reprezintă estimarea atenției lui Kevin asupra obiectului 2. Balta nu este deloc într-o poziție centrală, evidentă, din încăpere. Este în umbra mesei, așa că nu strălucește. Cu alte cuvinte, nu iese în evidență. Să presupunem că privirea lui Kevin nu se îndreaptă niciodată spre podea. E posibil ca el să fi procesat existența bălții în câmpul vizual periferic, dar la o primă privire mașinăria ar putea să estimeze în mod rezonabil o valoare mică a lui C_2 – și astfel o probabilitate scăzută ca Kevin să ocolească balta când se apropie de masă. Dacă nu e atent, o să calce fix în apă.

Cele două numere de până acum, C_1 și C_2, depind unul de altul. Sunt interconectate. Resursele de procesare ale lui Kevin sunt limitate, așa că, pe măsură ce C_1 crește, C_2 trebuie să scadă. Cu cât e mai atent la gogoașă, cu atât e mai puțin probabil să fie atent la baltă, și viceversa. Dacă atenția lui se concentrează asupra gogoșii când se apropie de masă, e probabil să calce în apă. Calcularea lui C în funcție de fiecare lucru din încăpere necesită, așadar, un model mai bogat, care să ia în considerare interacțiunile concurente dintre obiecte.

Mașinăria poate să calculeze și valoarea lui C_3 – estimarea gradului în care Kevin este atent la al treilea obiect, la telefon. Inițial, mașinăria calculează o valoare scăzută a lui C_3, dat fiind că telefonul este ascuns în umbră în partea din spate a încăperii și Kevin nu privește în acea direcție. Acum telefonul sună o singură dată, ieșind mai mult în evidență. Înregistrând această schimbare prin microfonul ascuns, mașinăria noastră de predicție poate calcula o creștere bruscă a lui C_3. În acel moment, atenția lui Kevin probabil se îndreaptă brusc spre telefon. Chiar dacă privirea lui ar fi îndreptată în altă parte, dată fiind intensitatea stimulului în acel moment, C_3 ar trebui să fie mare. Ca o consecință directă, C_1 și C_2 trebuie să scadă simultan. Cele trei numere se află constant într-un dans,

concurând între ele. În momentul în care telefonul sună, probabilitatea ca Kevin să ia gogoașa scade brusc la 0. Mai mult, atenția lui Kevin, atrasă de telefon când acesta sună, este cumva adezivă sau vâscoasă. Tinde să rămână concentrată asupra telefonului pentru cel puțin o jumătate de secundă după sunet și apoi va scădea treptat pe o curbă tipică atenției umane. Atenția nu trece instantaneu de la un obiect la altul. Pentru a face predicții optime despre comportamente, calculul lui C trebuie să țină seama de vâscozitatea atenției umane.

Punând laolaltă toate aceste date, Predictorul 5.000 trebuie să calculeze valorile în permanentă schimbare ale lui C_1 C_2 și C_3. Acest calcul se bazează pe informații de bază despre modul în care funcționează atenția umană. Mașinăria ia în calcul indicii cum ar fi gradul în care diferiți stimuli ies în evidență în mediu, numărul de stimuli din mediu, direcția în care privește Kevin, expresia lui facială și dinamica temporală a atenției umane. Calculând valorile fluctuante în timp ale lui C_1 C_2 și C_3, poate estima probabilitatea permanent schimbătoare ca Kevin să realizeze acțiuni în raport cu gogoașa, cu apa și cu telefonul.

Acest calcul al lui C este un model. Mecanismul de predicție construiește o descriere conform căreia Kevin are o proprietate – să-i spunem substanța C. Această substanță este invizibilă. Nu poate fi observată direct. Nu blochează și nu reflectă lumina. Este generată înăuntrul lui Kevin și se îndreaptă în exterior. Are tendința să se îndrepte în direcția privirii, deși nu urmează neapărat acea cale, dat fiind că uneori se poate îndrepta în direcții nonvizuale. Are contact cu obiecte specifice din mediu. Poate fi distribuită între obiecte, dar numai într-o manieră concurentă, astfel încât să poată fi îndreptată mai mult asupra unui singur obiect în detrimentul celorlalte. La fel ca apa care țâșnește dintr-un furtun, cu cât Kevin direcționează jetul spre un obiect anume, cu atât are mai puțină apă pentru altele.

Substanța C funcționează asemănător unui lichid

obișnuit.[0] Curge dintr-o sursă. Volumul său rămâne relativ același – în sensul că, cu cât curge mai mult aici, cu atât curge mai puțin acolo. Este și puțin vâscoasă – își schimbă direcția cu oarecare încetineală. În vreme ce se comportă ca un lichid invizibil, vâscos, are și o proprietate energetică, asemănătoare voinței, în sensul că îi dă celui în cauză puterea de-a acționa. Nu îl impulsionează direct pe Kevin și nici nu determină acțiuni specifice, însă prezența ei funcționează ca un generator care îi oferă lui Kevin capacitatea de-a face o alegere comportamentală.

Substanța C este un construct al mecanismului de predicție. Nu există în sine. Este o ilustrare a unor procese neuronale foarte reale, foarte complicate, care se desfășoară în capul lui Kevin. Aproximativ 80 de miliarde de neuroni din creierul lui îi controlează acțiunile, iar mecanismul de predicție îi atribuie ceva mult mai simplu, ceva schematic și foarte similar conștiinței metafizice.

Substanța C este un model simplificat al atenției corticale. Este o schemă a atenției.

Se pot întrevedea aplicații practice oarecum sinistre ale acestui tip de mecanism artificial de predicție. De exemplu, poate fi instalat în magazine pentru a prezice comportamentul oamenilor în timp ce fac cumpărături. Poate fi utilizat în scopuri de securitate sau pentru controlul mulțimilor, monitorizând atenția oamenilor într-un mod mult mai sofisticat decât simpla înregistrare a persoanelor prezente sau urmărirea direcției privirii cuiva. Poate fi utilizat în jocuri video pentru a ajuta personajele virtuale să prezică comportamentul jucătorilor. Atribuirea unei conștiințe invizibile, metafizice, unor agenți nu este o formă transcendentă de poezie; este un truc simplu, programabil, util pentru predicția comportamentului.

S-ar putea ca Predictorul 5.000 să nu aibă o rată de reușită prea mare. E dificil să prezici cât de cât corect comportamentul uman real. Kevin ar putea să se plimbe pur și simplu prin încăpere fără să facă nimic, mormăind în

[0] D.J. Acheson, Elementary Fluid Dynamics (Oxford, UK: Clarendon Press, 1990).

barbă, un comportament care nu este foarte predictibil. Reacțiile lui pot fi haotice sau pot fi determinate de factori ascunși, despre care nu știm nimic. Nu cred că noi, oamenii, ne pricepem prea bine la a prezice comportamentul de fiecare clipă al altcuiva și nu mă aștept ca inteligența artificială să fie mai bună la asta. Dar mașinăria nu trebuie să facă predicții absolut corecte. Dacă predicțiile pe care le face sunt *relativ* bune, având o probabilitate mai mare decât ceea ce este întâmplător, tot ne-ar oferi un avantaj util – același avantaj pe care oamenii îl obțin din predicțiile lor sociale imperfecte.

Să punem acum mașinăria într-o situație și mai dificilă. Kevin își poate îndrepta atenția spre o idee sau o emoție în aceeași măsură în care o poate direcționa spre gogoașă sau spre telefonul mobil. Dacă adăugăm ținte abstracte ale atenției, sarcina devine mai complexă, cu mai multe posibilități. Să ne imaginăm că Predictorul 5.000 își folosește microfonul pentru a asculta conversația pe care o poartă Kevin la telefon. Sarcina mașinăriei este să reconstruiască elementele spre care se îndreaptă atenția fluctuantă a lui Kevin, idei și teme de conversație, și apoi să prezică modul în care Kevin ar putea reacționa. Această sarcină de ascultare este, desigur, mult mai dificilă decât cea de vizualizare, dat fiind că informația pe care mașinăria trebuie să o asimileze este și mai redusă, și mai abstractă. Cuvintele lui Kevin și tonul vocii sale conțin toate informațiile disponibile. Dar principiile de bază rămân aceleași. Mașinăria identifică acele elemente, în acest caz idei abstracte, care îi pot atrage atenția lui Kevin; creează modelul unei substanțe C înăuntrul lui Kevin, o esență a conștiinței care se ocupă de aceste elemente; și folosește acest model pentru a prezice comportamentul verbal al lui Kevin.

Eu consider că oamenii au mecanisme de predicție care funcționează exact așa. Ne atribuim constant unii altora o subtilă substanță C, o conștiință a lucrurilor, un fel de forță sau de esență invizibilă și fluidă. Această forță este un fel de emanație, care se îndreaptă de obicei în aceeași direcție

în care privesc ochii. Ia contact cu obiectele din mediu. Are o energie – o calitate asemănătoare voinței, în sensul că le permite oamenilor să facă alegeri comportamentale și, astfel, să acționeze. Nu suntem neapărat conștienți că alcătuim acest model ciudat al atenției. O facem automat, având strania impresie că putem percepe cum ceilalți emană conștiință. Și facem toate astea pentru că evoluția a găsit o soluție pragmatică la o problemă importantă – prezicerea comportamentului celorlalți.

Am făcut de curând un experiment amuzant în laboratorul meu.[0] Un grup de voluntari trebuia să se uite la o imagine pe computer a unui cilindru de hârtie așezat vertical pe o masă. Participanților li s-a cerut să-și imagineze cum cilindrul se înclină din ce în ce mai mult și să estimeze unghiul critic la care avea să se prăbușească. Folosind săgețile de pe tastatură, participanții au marcat unghiul estimat în cazul unor cilindre cu înălțimi și diametre variabile. Uneori li s-a cerut să-și imagineze o înclinare spre stânga, alteori – spre dreapta. S-au descurcat destul de bine la această sarcină și au estimat unghiuri plauzibile din punct de vedere fizic.

Simultan, la fiecare probă am afișat pe ecran o imagine cu o față văzută din profil. Nu le-am explicat participanților de ce se afla acolo; pur și simplu era prezentă, fie în colțul din stânga, fie în colțul din dreapta al ecranului, privind direct spre cilindrul de hârtie. Când i-am întrebat apoi pe participanți de ce credeau că am inclus imaginea unei fețe în experiment, au dat diverse răspunsuri, dar niciunul nu a ghicit motivul real. Cei mai mulți au crezut că fața nu le-a influențat deloc performanța. Și totuși le-a afectat subtil judecata, de parcă ar fi perceput o undă de energie care pornea de la ochii din imagine și influența unghiul de

[0] A. Guterstam, H.H. Kean, T.W. Webb, F.S. Kean și M.S.A. Graziano, „An Implicit Model of Other People’s Visual Attention as an Invisible, Force-Carrying Beam Projecting From the Eyes”, Proceedings of the National Academy of Sciences of the United States of America 116 (1) (2019): 328 – 33

înclinare al cilindrului de hârtie. Când cilindrul se înclina în direcția în care era poziționată fața, privirea părea să îl susțină, și oamenii considerau că se putea înclina mai mult înainte să cadă. Când cilindrul se înclina în direcția opusă feței, acea undă părea cumva să îl împingă, și oamenii considerau că ar cădea mai devreme, la un unghi mai ascuțit. Însă, când fața era legată la ochi, acest efect dispărea. Unghiul la care cilindrul s-ar fi prăbușit era același indiferent dacă tubul de hârtie cădea în direcția feței sau în direcția opusă, ca și cum faptul că ochii erau acoperiți anula unda și puterea ei.

Am putut chiar să calculăm forța fictivă exercitată de undă, dat fiind că știam efectul pe care îl avea asupra înclinării cilindrului ușor de hârtie. Magnitudinea era foarte mică, însă îndeajuns de puternică pentru a înclina subtil cilindrul: aproximativ o sutime de newton. Asta înseamnă ceva mai puțin decât forța exercitată de o stafidă pe care o ții în palmă în condițiile gravitației de pe Pământ. Desigur, era de așteptat ca această forță să fie subtilă. Dacă oamenii ar percepe o forță intensă emanând din ochii celorlalți, ca o rafală de vânt care ar putea răsturna un coș de gunoi, atunci conflictul dintre percepție și realitate ar deveni evident, iar percepțiile greșite le-ar pune grav în pericol supraviețuirea. Adevărul e că această forță fictivă este atât de subtilă, încât de-abia poate fi măsurată.

Participanții la studiul nostru nu erau conștienți de această tendință pe care o prezentau. În chestionarul pe care l-au completat ulterior, susțineau că judecățile lor cu privire la unghiul de înclinare nu au fost deloc influențate de fața de pe ecran. Pentru ca lucrurile să fie mai clare, i-am întrebat cum credeau ei că funcționează văzul: implică ceva ce iese din ochi, sau ceva ce intră? Aproximativ 5% se aflau în eroare și au spus că e ceva ce iese din ochi. Ceilalți au indicat în mod corect că văzul funcționează datorită luminii care pătrunde în ochi. Și, cu toate astea, în mod implicit, toți păreau să se raporteze la ochi ca și cum din aceștia ar emana o substanță invizibilă care interacționează cu lumea fizică. Nu se puteau abține. În interpretarea mea,

aveam de-a face cu substanța C. Ce observam era acțiunea unui model simplificat al atenției.

Ideea că atenția vizuală poate emana din ochi și poate influența fizic obiectele din lumea exterioară nu este deloc nouă. E una dintre cele mai durabile credințe din înțelepciunea populară. Se numește teoria extramisiei văzului și datează cel puțin de la filosofii greci din Antichitate.[0] Această teorie a bătut cale lungă și a avut susținători iluștri, ca Platon sau, mai târziu, medicul roman Galen. În secolul al IX-lea d.Hr., savantul arab Ibn al-Haytham a descoperit în sfârșit legile corecte ale opticii și a declarat că extramisia nu era întemeiată. Lumina pătrunde în ochi în unghiuri drepte pentru a forma o imagine.

În ciuda acestui adevăr științific definitiv, credința în undele oculare a stăruit. La o mie de ani după al-Haytham, credința în „deochi” se regăsește încă în multe culturi, alături de un comerț prosper cu amulete care te protejează.[0] în cultura noastră, Superman are vedere cu raze X, care îi țâșnesc din ochi și împrăștie flăcări în jur. Aproape toată lumea a avut la un moment dat senzația neplăcută că este privit din spate, ca și cum ar fi simțit pe ceafă o undă de energie. În 1898, psihologul Edward Titchener a considerat că această credință era într-atât de răspândită încât merita să fie testată.[0] Prin experimente controlate a descoperit că oamenii nu pot să-și simtă unii altora privirea, oricât de puternică ar fi această impresie psihologică.

Credința în undele oculare este într-adevăr seducătoare din punct de vedere psihologic. Când aveam vreo cinci ani, stăteam cu tatăl meu pe scările verandei și ne uitam la stele. M-a întrebat cum de puteam să văd obiecte atât de îndepărtate. Retrospectiv, cred că voia să zică ceva despre

---

[0] C.G. Gross, „The Fire that Comes From the Eye”, The Neuroscientist 5 (1999): 58 – 64.

[0] A. Dundes, The Evil Eye: A Folklore Casebook (New York: Garland Press, 1981).

[0] E.B. Titchner, „The Feelingof Being Stared At”, Science 8 (1898): 895 – 897.

lumina care călătorește milioane de ani, dar mi-e teamă că am schimbat subiectul. Am început să-i explic că, atunci când mă uit în sus, o chestie vizuală îmi iese din ochi și se duce spre cer. Am văzut reacția stânjenită a omului de știință. N-a precupețit niciun moment să-mi explice cu blândețe fundamentele opticii. Da, îmi amintesc exact momentul în care am trecut de la o teorie a vederii bazate pe extramisie la o teorie bazată pe intromisie. Poate că acela a fost un moment definitoriu pentru viața mea de om de știință, deși se pare că toată lumea trece prin această schimbare. Credința că vederea implică ceva ce iese din ochi este atât de intuitivă – atât de intrinsecă modului în care înțelegem cum funcționează văzul –, încât reprezintă convingerea pe care o au îndeobște copiii, după cum a descoperit psihologul Jean Piaget.[0] Abia mai târziu învățăm explicația corectă din punct de vedere științific.

O serie de studii realizate în anii 1990 au descoperit că majoritatea elevilor din SUA credeau în teoria greșită a extramisiei. În acele studii, până la 60% dintre elevi susțineau că vederea funcționează nu pentru că ceva pătrunde în ochi, ci pentru că ceva iese din ochi. Trebuie totuși să adaug că sunt sceptic în privința acestor rezultate extreme. Nu știu dacă felul în care au fost puse întrebările a influențat răspunsul, sau dacă știința pedagogică s-a îmbunătățit radical în ultimii 30 de ani, ori dacă participanții nu râdeau poate de cercetători. Descoperirile noastre contrazic aceste date care arată că un procent atât de mare de oameni au credințe greșite, iar noi am studiat o populație destul de variată în privința nivelului de educație și de venituri. Să crezi în teoria undelor oculare pare să fie ceva rar în zilele noastre, majoritatea adulților înțelegând fundamentele opticii.

Cu toate astea, în ciuda cunoașterii științifice, se pare că avem încă tendința să credem în extramisie. Pentru mine, cea mai interesantă descoperire din studiul nostru cu cilindrul de hârtie a fost opoziția uimitoare dintre

[0] J. Piaget, The Child’s Conception of the World, trad. De J. Tomlinson și A. Tomlinson (Totowa, NJ: Little, Adams, 1979).

cunoașterea intelectuală a oamenilor și intuiția lor neexaminată. Nu a contat că știau cum stau lucrurile de fapt; la nivel implicit, considerau încă vederea ca fiind ceva ce iese din ochi. Presupun că această intuiție nu se rezumă la atenția vizuală și că se aplică oricărui tip de atenție. În cultura noastră circulă nenumărate credințe despre modul în care concentrarea intensă a atenției poate influența lucrurile, atingând oameni sau obiecte. Mă întreb câți copii, după ce au văzut pentru prima oară *Războiul stelelor*, au încercat să se concentreze asupra unui creion pentru a vedea dacă pot să-l facă să se miște. Eu unul am încercat. Clasica undă vizuală este, din punctul meu de vedere, mai degrabă o undă *mintală* care adesea iese din ochi, dar nu se limitează neapărat la văz.

Motivul pentru care aceste convingeri au o asemenea pondere culturală poate fi faptul că sunt legate de un model implicit, profund, automat, cu care toți ne naștem și care a evoluat pe parcursul a milioane de ani. Acest model ne ajută să urmărim atenția celorlalți într-o manieră eficientă, schematică, astfel încât să le putem prezice comportamentul. Chiar dacă din punct de vedere intelectual știm mai bine, nu ne putem abține să nu generăm această intuiție. Nu putem să nu ținem seama în mod implicit de undele atenției celorlalți.

Cred că această intuiție este și motivul pentru care conștiința este atât de greu de studiat științific. Multe teorii ale conștiinței sunt blocate într-o presupunere neștiințifică, în esență mistică. Dacă încercăm să înțelegem cum generează creierul o eterică substanță C, nu vom ajunge niciodată la o cunoaștere științifică. Oamenii privesc înăuntru și accesează modele care nu au evoluat pentru a fi descrieri corecte din punct de vedere științific ale lumii fizice. Aceste modele au fost create și cizelate în mai bine de 300 de milioane de ani de evoluție, optimizate pentru a fi utile în moduri specifice, pragmatice. Sunt schematizate și dezgolite de detalii inutile. Suntem tentați să înțelegem literal conținuturile acestor modele și apoi să ne lansăm într-o căutare științifică a undelor vizuale sau a unei energii

lăuntrice, invizibile, ori a unei esențe subiective. Încercăm să înțelegem modul în care creierul poate genera substanța C. Mă îndoiesc că vom găsi vreodată o soluție tehnică pentru construirea unui câmp mintal magic. Însă cu siguranță are sens construirea unei mașinării de predicție comportamentală care așază tocmai acest tip de construct simplificat în centrul operațiilor pe care le realizează.

Teoria schemei atenției, alături de alte teorii mecaniciste, e acuzată uneori că ar devaloriza conștiința, că ar face din ea doar o iluzie (și voi avea mai multe de spus despre propunerea „iluzionismului" mai târziu în acest capitol). Dar în teoria schemei atenției, întrucât conștiința poate fi înțeleasă mecanicist, pot fi înțelese și obiectivele ei specifice, pragmatice. În loc să anuleze conștiința, teoria o pune în centrul capacităților noastre. Este un model intern străvechi, foarte simplificat, cizelat de evoluție pentru a servi două funcții principale, după cum am arătat în ultimele câteva capitole. Prima sa funcție a fost probabil aceea de model al sinelui, cu scopul de a monitoriza, de a prezice și de a controla atenția proprie. A doua funcție ar putea fi aceea de catalizator al cogniției sociale, care ne permite să intuim stările atenționale ale celorlalți și să le prezicem astfel comportamentul. Prin urmare, consider nu numai că conștiința poate fi înțeleasă științific sau că va putea fi în cele din urmă construită de tehnicieni, ci și că este un instrument de o extraordinară putere și importanță practică.

6
Yoda și Darth:
Cum putem descoperi conștiința în creier?

Prima întrebare pe care o pune oricine despre conștiință și despre creierul uman este: „Unde se află? În care parte a creierului e situată conștiința?"

Desigur, să știi unde e localizată o funcție cum e conștiința nu explică ce este și cum apare. Să presupunem că nu știi nimic despre cum e stocată memoria într-un computer. Un tehnician ar putea să scoată carcasa, să indice un obiect dinăuntru și să spună: „Ăsta e cipul de memorie”. Cu această informație nu știi mult mai multe decât știai înainte, dar măcar știi de unde să începi. Știi cărui obiect să i te adresezi pentru răspunsuri. Să sperăm că și localizarea conștiinței în creier va fi utilă în același mod.

Structura cerebrală indicată cel mai frecvent ca loc al conștiinței este cortexul cerebral, partea creierului care s-a dezvoltat cel mai mult pe parcursul evoluției oamenilor.[0]1 A doua este talamusul.[0] Cum am menționat înainte, talamusul are o conexiune strânsă cu cortexul și cele două fac schimb constant de informații, fiecare parte a talamusului comunicând în special cu o parte anume a cortexului.[0]

O a treia ipoteză este o misterioasă structură cerebrală numită claustru, un strat subțire de celule situat imediat sub cortex, pe cele două părți ale creierului, aproape de urechi.[0] În 1987, când eram un student naiv și fără experiență și când am început să lucrez pentru prima dată într-un laborator de neuroștiințe, l-am întrebat pe profesor dacă pot studia claustrul, mai ales pentru că nimeni nu părea să știe nimic despre el. M-a lăsat să încerc, dar

---

[0] J.H. Kaas, „The Evolution of Brains From Early Mammals to Humans”, Wiley Interdisciplinary Review of Cognitive Science 4 (2013): 33 – 45.

[0] J.E. Bogen, „Some Neurophysiologic Aspects of Consciousness”, Seminars în Neurology 17 (1997): 95 – 103; G.M. Edelman, J.A. Gaily și B. J. Baars, „Biology of Consciousness”, Frontiers în Psychology 2 (2011): 4; L.M. Ward, „The Thalamic Dynamic Core Theory of Conscious Experience”, Consciousness and Cognition 20 (2011): 464 – 486.

[0] E.G. Jones, The Thalamus (New York: Springer, 1985).

[0] F.C. Crick, C. Koch, „What is the Function of the Claustrum?”, Philosophical Transactions of the Royal Society of London, B, Biological Sciences 360 (2005): 1271 – 1279; Y. Goll, G. Atlan și A. Citri, „Attention: The Claustrum”, Trends în Neurosciences 38 (2015): 486 – 495; Z.K. Mohamad, B. Fabrice, B. Abdelrahman și P. Fabienne, „Electrical Stimulation of a Small Brain Area Reversibly Disrupts Consciousness”, Epilepsy and Behavior 37 (2014): 32 – 35.

claustrul este o zonă cerebrală atât de subțire, încât era dificil să obțin vreun rezultat palpabil, așa că am trecut la alte proiecte. Am în continuare o slăbiciune pentru claustru și citesc cu nerăbdare noile informații științifice despre el. E conectat la cortex printr-o rețea de procesare nu foarte diferită de cea dintre talamus și cortex. Dar, chiar dacă au fost descoperite conexiunile, funcția claustrului rămâne necunoscută. Același lucru se poate spune despre multe structuri cerebrale. Faptul că cea mai mare parte a creierului este un mister necartografiat reprezintă una dintre bucuriile pe care ți le oferă studiul neuroștiinței.

Aceste trei structuri ale creierului sunt atât de strâns legate între ele încât formează un singur sistem. Ar fi corect să spunem că conștiința umană este probabil o funcție a cortexului*, unde asteriscul înseamnă „plus alte structuri, ca talamusul și claustrul, fără de care cortexul nu funcționează”. Întrucât cortexul se întinde atât de convenabil pe toată suprafața creierului, e mult mai simplu de studiat decât alte structuri și, prin urmare, e obiectul celor mai multe eforturi experimentale. Se încearcă acum să se descopere dacă toate părțile cortexului contribuie la conștiință sau dacă unele regiuni au un rol mai important decât altele. Oricare ar fi părțile cortexului care contribuie cel mai mult la conștiință, sunt probabil implicate și regiunile interconectate ale talamusului, claustrului și ale multor alte structuri cerebrale.

Specialiștii în neuroștiințe se fac uneori vinovați de ceea ce a fost numit zonologie – colorează cu roșu o zonă de pe suprafața creierului, îi atribuie o funcție și se poartă de parcă ar fi aflat ceva nou. Cortexul este împărțit în mod tradițional în regiuni de dimensiunea unor timbre poștale, fiecare cu proprietățile sale. Dar, pe măsură ce au aflat mai multe, oamenii de știință au ajuns să înțeleagă faptul că este alcătuit din rețele extinse, nu din zone izolate. Ariile cerebrale vecine deschid conexiuni funcționale unele cu celelalte, așa cum țările deschid canale diplomatice, iar alianțele dintre zone se schimbă în funcție de starea mintală a persoanei și de sarcina mintală specifică pe care

o realizează.

În acest capitol voi prezenta o parte din experimentele care s-au făcut pentru descoperirea rețelelor neuronale specifice din cortex care sunt responsabile pentru conștiință. Nu pretind nicidecum că voi acoperi toate sau măcar cea mai mare parte a excelentelor eforturi experimentale care s-au făcut până acum. Am selectat doar câteva exemple pentru a ilustra dificultățile conceptuale profunde, uneori ascunse, inerente studierii conștiinței.

Specialistul în neuroștiințe Sabine Kastner lucrează cu un etaj mai jos în clădirea în care lucrez și eu la Universitatea Princeton. E un expert de renume mondial în domeniul atenției vizuale. Date fiind colaborările noastre, nu e deloc de mirare că am dezvoltat o teorie a conștiinței centrată pe noile descoperiri în materie de știința atenției. Apropo, dr. Kastner nu este doar o colaboratoare – este soția mea. Am avut multe conversații la cină despre conștiință și despre creier. Pe masa noastră stă, cu rol decorativ, un model tridimensional al creierului fiului nostru de 11 ani, construit după o tomografie. Soția mea și cu mine ne entuziasmăm uneori vorbind despre convoluțiile acestuia, în timp ce fiul meu își dă ochii peste cap și încearcă să readucă conversația la propriile lui interese (fauna din cretacicul târziu).

În urmă cu câțiva ani, ca parte a unei conferințe academice obișnuite, soția mea a făcut o demonstrație pentru public. A împărțit ochelari 3D care separau imaginile văzute de ochiul drept de cele văzute de ochiul stâng. Apoi a afișat o imagine cu Yoda creată astfel încât să fie percepută cu ochiul drept, și o imagine cu Darth Vader creată astfel încât să fie percepută cu ochiul stâng. Din perspectiva audienței, Yoda și Vader nu puteau fi văzuți în același timp. Fenomenul se numește rivalitate binoculară. Mai întâi îl vezi pe Yoda. Apoi, după câteva secunde, Yoda începe să se estompeze, dispare, și apare Darth. După câteva secunde, Darth se estompează, și Yoda triumfă. Și tot așa, într-o luptă în stil binocular între yin-ul și yang-ul

Forței.

Rivalitatea binoculară a fost unul dintre primele instrumente pe care le-au folosit cercetătorii în încercarea de a localiza conștiința în creier, și este în continuare unul dintre cele mai populare.[0] Să presupunem că pui pe cineva într-un scaner RMN pentru a-i măsura activitatea cerebrală și că descoperi o zonă din cortexul cerebral în care neuronii reacționează intens la Yoda. Poate că procesează selectiv imaginile verzi. Dacă îl arăți doar pe Yoda, neuronii de acolo se activează. Când îl arăți pe Darth Vader, neuronii devin inactivi. Acum, să presupunem că afișezi imagini duale, competitive – Yoda într-un ochi, Darth în celălalt – și măsori activitatea acelor neuroni care receptează culoarea verde.

Întrucât Yoda este prezent constant în ochiul drept, neuronii pot fi activi continuu. Dacă e așa, atunci acești neuroni probabil nu sunt implicați în conștiința vizuală, ci au de-a face cu procesarea brută a imaginii afișate. Dar să presupunem că, în schimb, acei neuroni sunt activi când imaginea lui Yoda ajunge în conștiința subiectului și apoi rămân inactivi când subiectul spune că Yoda s-a estompat și că Darth a ajuns în conștiința vizuală. În acest caz, activitatea neuronilor se corelează cu conștientizarea lui Yoda. Poate că acei neuroni, când se activează și îl procesează pe Yoda, *fac* ca persoana să fie conștientă de

[0] R. Blake, J. Brascamp și DJ. Heeger, „Can Binocular Rivalry Reveal Neural Correlates of Consciousness?”, Philosophical Transactions of the Royal Society of London, B, Biological Sciences 369 (2014): 20130211; R. Blake și N.K. Logothetis, „Visual Competition”, Nature Reviews Neuroscience 3 (2002): 13 – 21; D.A. Leopold și N.K. Logothetis,» Activity changes în Early Visual Cortex Reflect Monkeys’ Percepts During Binocular Rivalry”, Nature 379 (1996): 549 – 553; B.A. Metzger, K.E. Mathewson, E. Tapia, M. Fabiani, G. Gratton și D.M. Beck, „Regulating the Access to Awareness: Brain Activity Related to Probe-related and Spontaneous Reversals în Binocular Rivalry”, Journal of Cognitive Neuroscience 29 (2017): 1089 – 1102; K. Sandberg, B. Bahrami, R. Kanai, G.R. Barnes, M. Overgaard și G. Rees, „Early Visual Responses Predict Conscious Face Perception Within and Between Subjects During Binocular Rivalry”, Journal of Cognitive Neuroscience 25 (2013): 969 – 985; F. Tong, M. Meng și R. Blake, „Neural Bases of Binocular Rivalry”, Trends în Cognitive Sciences 10 (2006): 502 – 511.

Yoda. Similar, trebuie să existe o altă serie de neuroni care, când se activează, fac ca persoana să fie conștientă de Darth Vader. Cel puțin la suprafață, experimentul pare să aibă sens.

Această metodă a fost folosită pentru prima dată în anii 1990 în cercetările asupra creierului maimuțelor și este încă utilizată în cercetările asupra creierului uman, în căutarea acelor părți ale sistemului vizual care se corelează cu și poate chiar generează conștiința.[0] Inițial, cercetătorii au presupus că regiunile corticale de la baza ierarhiei procesării, prin care informațiile intră în sistem, cum ar fi zona V1, vor răspunde la imaginile de la ambii ochi într-o manieră constantă, neindicând vreo relație cu conștiința. La urma urmei, acele zone par să proceseze informația la un nivel prea simplu pentru a crea conștiința. Prin opoziție, regiunile corticale de la vârful ierarhiei ar putea răspunde trecând de la o imagine la alta, indicând astfel modul în care se deplasează conștiința. Aceste zone procesează informația vizuală într-un mod mai bogat, mai complex, holistic; prin urmare, ele trebuie să fie cele care generează conștiința vizuală. Undeva între aceste două extreme era de așteptat să se producă un salt. Pe măsură ce informația vizuală avansează în ierarhia de procesare, devenind din ce în ce mai elaborată și mai complexă, ar trebui să ajungă la un prag care să o trimită în conștiință.

Acest rezultat simplu nu s-a materializat. În schimb, efectele se distribuie în întreg sistemul vizual.[0] Nu doar sistemul cortical, ci și un stadiu anterior de procesare, regiunea talamusului care conectează ochiul la cortex, prezintă o corelație parțială cu percepția conștientă.[0]

---

[0] R. Blake și N.K. Logothetis, „Visual Competition”, Nature Reviews Neuroscience 3 (2002): 13 – 21; D.A. Leopold și N.K. Logothetis, „Activity changes în Early Visual Cortex Reflect Monkeys’ Percepts During Binocular Rivalry”, Nature 379 (1996): 549 – 553.

[0] R. Blake și N.K. Logothetis, „Visual Competition”, Nature Reviews Neuroscience 3 (2002): 13 – 21; F. Tong, M. Meng și R. Blake, „Neural Bases of Binocular Rivalry”, Trends în Cognitive Sciences 10 (2006): 502 – 511.

[0] K. Wunderlich, K.A. Schneider și S. Kastner, „Neural Correlates of

Activitatea acesteia alternează între ochiul drept și cel stâng pe măsură ce subiectul devine conștient de o imagine sau de alta. Cu cât privim mai sus în ierarhie, cu atât neuronii tind să se coreleze cu percepția vizuală conștientă, dar nu pare să existe o limită clară între părțile conștiente și cele neconștiente.

Este tentantă concluzia potrivit căreia conștiința apare treptat în sistemul vizual, acumulându-se pe măsură ce informațiile urcă în ierarhia procesării. Conform acestei perspective, creierul nu are un loc anume care generează conștiința. În schimb, conștiința este o proprietate a întregului sistem. Să fie această interpretare corectă, sau experimentul rivalității binoculare, în ciuda eleganței sale, ascunde niște demoni conceptuali?

Să presupunem că sarcina noastră este să înțelegem modul în care oamenii văd culorile. Într-un fel, culoarea se aseamănă conștiinței. Nu ne gândim la culoare ca având în mod intrinsec substanță sau greutate, dar ea există și poate fi atașată obiectelor din lumea înconjurătoare. În consecință, voi adopta o perspectivă cu desăvârșire greșită asupra culorii – dar logica ei este similară celei pe care oamenii o aplică uneori conștiinței.

Imaginează-ți că te uiți la un măr. O întrebare tradițională în neuroștiințe ar fi: „Unde în creierul tău se duce informația despre măr pentru a genera o experiență vizuală conștientă?” Să punem o întrebare paralelă: „Unde în creier se duce informația despre măr pentru a genera culoarea roșie?”

Sistemul vizual procesează forma obiectelor. V1, zona corticală situată la baza ierarhiei de procesare, se ocupă de aspecte simple ale formei, cum ar fi înclinația unui mic segment de linie aflat la granița unei forme. Mai sus în ierarhie, în regiunile corticale de ordin superior, stau alături aspecte mai complexe și mai integrate ale formei. Să formulăm o teorie a culorii – știu că e greșită, dar o facem

Binocular Rivalry în the Human Lateral Geniculate Nucleus”, Nature Neuroscience 8 (2005): 1595 – 1602.

de dragul demonstrației – în care culoarea ia naștere din modul în care creierul procesează forma. Conform acestei teorii, când informația vizuală ajunge pentru prima oară în sistem, nu conține culoare, fiind codificată doar în alb și negru. Acum să presupunem că undeva în secvența de procesare culoarea începe să apară pe măsură ce informația despre formă este procesată într-o manieră mai complexă. Până când informația ajunge la nivelurile superioare ale ierarhiei vizuale, culoarea este pe deplin generată din informația despre formă. În această teorie ciudată, culoarea este o esență generată de creier, un produs secundar al modului în care este procesată informația. Să o numim teoria culorii emergente.

Acum, că am construit această teorie incorectă despre cum ia naștere culoarea, următorul pas este să descoperim ce părți ale cortexului generează culoarea. Pentru asta, vom folosi rivalitatea binoculară în cazul unui subiect uman în timp ce-i măsurăm activitatea cerebrală prin RMN. Vom aplica exact aceeași logică a experimentului privitor la conștiință pe care l-am prezentat mai înainte. Cele două imagini vor fi un pătrat roșu și un cerc roșu. Pătratul este proiectat în ochiul drept, cercul, în ochiul stâng. Participantul nostru are acum o experiență a rivalității – vede un pătrat roșu timp de câteva secunde, după care pătratul se estompează și vede un cerc roșu, și tot așa. Proprietatea culorii roșii este atașată mai întâi pătratului, apoi cercului, alternativ. Într-un moment, informațiile despre pătrat trebuie să fi ajuns în acele regiuni cerebrale unde generează esența roșului. În următorul moment, informațiile despre cerc trebuie să fi pătruns către regiunile cerebrale unde pot genera roșul. Dar unde sunt acele regiuni?

Să presupunem că descoperim o regiune a cortexului care se activează când subiectul nostru raportează că vede un pătrat roșu, dar devine inertă când spune că vede cercul roșu. Acea regiune cerebrală este, pesemne, un generator de culoare. Neuronii ei procesează informațiile despre pătrat și, când sunt activați, acțiunea lor trebuie să

genereze și roșul din aceste informații. Probabil că o altă grupare de neuroni, activați când subiectul spune că vede un cerc roșu, servesc o funcție similară de generare a roșului în cazul cercului.

În timp ce facem experimentul descoperim că semnele distincte ale rivalității dintre pătrat și cerc sunt răspândite pe toată suprafața cortexului vizual, de la imput la cele mai înalte niveluri ale ierarhiei de procesare. Informația despre formă pare să creeze culoarea, oriunde s-ar afla în sistem. Pe baza acestui rezultat ajungem la concluzia că culoarea este o proprietate diseminată, holistică, neputând fi indicată o singură regiune cerebrală care o generează. Ia naștere din toate părțile sistemului.

Înțeleg cât de absurdă e teoria culorii emergente și înțeleg cât de slab e experimentul rivalității binoculare pe care l-am propus. Nu le elaborez aici pentru a râde de abordările anterioare ale conștiinței, ci ca să arăt cât de ușor e să cazi pradă unei linii greșite de gândire. La urma urmei, înainte de înțelegerea științifică a culorii, teoria culorii emergente ar fi putut părea plauzibilă. Au fost propuse teorii și mai ciudate. Astăzi știm multe despre culoare, dar este încă ușor să te lași păcălit de o teorie echivalentă a conștiinței emergente. Expun în continuare câteva dintre problemele logice ale acestei linii de gândire.

În primul rând, culoarea nu e o proprietate emergentă. Într-un sens, da, vederea culorii „emerge” din rețele de neuroni. Același lucru se poate spune despre orice însușire pe care o procesează creierul. Dar, pentru ca subiectul nostru să spună că pătratul este roșu, ceva din creierul lui trebuie să fi procesat informații specifice despre culoarea roșie. Poți să construiești un sistem vizual spectaculos de complicat și integrat și holistic; un sistem care să proceseze forma în sus și în jos și în orice direcție; dar pot să-ți garantez că, dacă nu montezi partea care procesează în mod specific informația despre culoare, sistemul vizual nu va fi sensibil la culoare. Culoarea va fi irelevantă pentru el. Nu e corectă ideea că, dacă informația despre pătrat este intens procesată sau este extrem de complicată ori e

procesată într-o anumită regiune cerebrală superioară, atunci creierul va începe să genereze culoare. Dacă culoarea este o esență eterică generată de procesarea formei, dacă nu există nicio informație specifică despre culoare care să fie procesată ca orice alt aspect al stimulului, atunci cum poate persoana să raporteze verbal că vede culoarea? La urma urmei, pentru a face o afirmație, creierul are nevoie de informația necesară care să susțină afirmația. Pentru a înțelege modul în care oamenii văd un pătrat roșu, nu e indicat să ne punem ochelari de cal și să ne uităm la regiunile cerebrale care procesează informația despre forma pătratului, sperând să descoperim ce zone generează și culoare. În schimb, ar trebui să căutăm circuitele în care sunt procesate informațiile despre culoare.

La fel, pentru a înțelege conștiința vizuală, nu trebuie să ne uităm la ariile cerebrale în care este procesată informația, sperând să descoperim care dintre acestea generează esența conștiinței. Ar trebui, în schimb, să căutăm rețelele care construiesc informațiile specifice despre conștiință.

A doua eroare din experimentul nostru fictiv este că ambii Stimuli au culoare și că e *aceeași* culoare. Indiferent dacă subiectul percepe un pătrat roșu sau un cerc roșu, tot roșu procesează. Orice parte a sistemului vizual care procesează culoarea va reacționa în același mod în ambele circumstanțe; nu se va activa sau dezactiva în funcție de trecerea de la pătrat la cerc. Dacă există regiuni corticale specializate în informații despre culoare – și se pare că într-adevăr există multe arii cerebrale care tocmai asta fac[0] –, atunci experimentul nostru nu ar avea cum să le găsească. În mod asemănător, când studiem conștiința, indiferent

[0] M.S. Beauchamp, J.V. Haxby, f. E. Jennings și E.A. Deyoe, „An FMRI Version of the Farnsworth-Munsell 100-Hue Test Reveals Multiple Color – Selective Areas în Human Ventral Occipitotemporal Cortex”, Cerebral Cortex 9 (1999): 257 – 263; B.R. Conway, „Color Signals Through Dorsal and Ventral Visual Pathways”, Visual Neuroscience 31 (2014): 197 – 209.

dacă subiectul este conștient de imaginea din ochiul stâng sau de imaginea din ochiul drept, e la fel de conștient de o imagine în ambele cazuri. Orice regiune cerebrală care procesează informații despre prezența conștiinței ar trebui să fie la fel de activă în ambele situații. Nu s-ar activa sau dezactiva în funcție de imaginea care domină.

Experimentul ar rata acele arii ale creierului care procesează constructul de conștiință.

A treia eroare a experimentului vine din neînțelegerea sursei rivalității binoculare. Rivalitatea binoculară este generată de competiție.[0] Semnalele de la cei doi ochi încep să interacționeze aproape imediat ce ajung în creier. Această interacțiune se amplifică pe măsură ce informația pătrunde mai profund în sistemul vizual. Dacă imaginile transmise de fiecare ochi sunt suficient de diferite, ca în cazul lui Yoda și Darth sau ca în exemplul cu pătratul și cercul, atunci sistemul nu le poate contopi în mod util și cele două serii de semnale ajung să se inhibe una pe cealaltă.

Semnalul care provine de la ochiul drept și semnalul care provine de la ochiul stâng sunt ca doi luptători la fel de puternici. Domină când unul, când celălalt. Niciunul nu ajunge la o victorie decisivă. Orice semnal care câștigă temporar domină sistemul cortical, inhibând celălalt semnal. Dacă descoperi efecte ale rivalității în talamus, sau în V1, sau la orice alt nivel al ierarhiei vizuale, niciunul dintre acestea nu indică faptul că însăși conștiința este prezentă sau procesată în acea regiune cerebrală. Ce arată însă rezultatele este că cele două semnale vizuale care provin de la fiecare ochi sunt într-o competiție constantă care se desfășoară în întregul sistem vizual.12 Oricare din imagini câștigă această competiție ajunge în cele din urmă

[0] R. Blake și N.K. Logothetis, „Visual Competition”, Nature Reviews Neuroscience 3 (2002): 13 – 21; H.H. Li, J. Rankin, J. Rinzel, M. Carrasco și D.J. Heeger, „Attention Model of Binocular Rivalry”, Proceedings of the National Academy of Sciences U.S.A. 114 (2017): E6192-E6201; F. Tong, M. Meng și R. Blake, „Neural Bases of Binocular Rivalry”, Trends în Cognitive Sciences 10 (2006): 502 – 511.

să fie atașată conștiinței; nu aflăm însă mai multe despre sistemul cerebral care construiește conștiința sau despre cum este aceasta atașată informației vizuale.

Utilizarea rivalității binoculare pentru studierea conștiinței este o bună ilustrare a modului în care un experiment poate părea elegant la început, dar, odată ce presupunerile sunt examinate mai în profunzime, se transformă într-o mlaștină științifică. Erorile de logică pot fi evidente în ceea ce privește percepția culorii, dar, când vorbim despre conștiință, este surprinzător de simplu și comod să cădem în aceeași greșeală. Rivalitatea binoculară reprezintă în continuare un fenomen minunat, care merită studiat la nivelul neuronilor și al regiunilor cerebrale. Mulți oameni de știință, inclusiv soția mea, continuă să o exploreze pentru relevanța ei intrinsecă și pentru ceea ce spune despre competiția vizuală. Dar se dovedește a fi un instrument nu tocmai potrivit pentru a descoperi unde este localizată conștiința. Ne trebuie alt experiment.

Voi propune o abordare experimentală mult mai puțin ingenioasă decât experimentul cu rivalitatea binoculară. E rudimentară și aproximativă, dar e pragmatică. Să începem prin a căuta rețelele cerebrale de procesare a culorii și să aplicăm apoi această logică descoperirii rețelelor de procesare a conștiinței.

Dacă aș vrea să găsesc centrele de procesare a culorii din cortex, aș pune subiectul nostru într-un aparat RMN și i-aș arăta o serie de poze alb-negru. Aș intercala uneori câte una color. Dacă o parte a cortexului vizual s-ar activa mai intens la vederea culorii, atunci aș ști că trebuie să studiez mai îndeaproape acea regiune.

Acest experiment este încurajator de simplu. Ca regulă generală, cu cât un experiment are mai puține componente, cu atât sunt mai mari șansele de reușită. Dar, în același timp, experimentul propus nu este perfect. În primul rând, alb și negru sunt și ele culori, procesate ca orice altă culoare. Experimentul compară imagini care au o gamă mai *redusă* de culori cu imagini cu o gamă mai *largă*

de culori în speranța că diferența este suficientă pentru a activa zonele de procesare a culorii. O a doua capcană este că experimentul ar putea să activeze și alte regiuni cerebrale, care nu au nimic de-a face cu culoarea. De exemplu, s-ar putea ca o regiune cerebrală să fie implicată în controlul vigilenței, activându-se când o imagine colorată este afișată în mod neașteptat după un minut de imagini plictisitoare în alb-negru. Dar, în ciuda imperfecțiunilor sale, m-aș încrede în această metodă ca punct inițial de căutare a răspunsului.

Metoda a mai fost folosită și reprezintă o modalitate excelentă de identificare a principalelor rețele de procesare a culorii.[0] În sine, experimentul este doar un început, dar rezultatele sale au fost confirmate și de alte metode. De exemplu, dacă studiezi oameni care au suferit vătămări ale acestor regiuni provocate de atacuri cerebrale, vei descoperi că acești pacienți și-au pierdut capacitatea de percepție a culorii.[0] Văd lumea în gri și și-au pierdut chiar capacitatea de a-și aminti sau de a-nțelege pe deplin ce este culoarea.

Dat fiind că mai multe experimente simple care abordează problema din unghiuri diferite converg către aceleași rezultate, descoperirile sunt din ce în ce mai sigure. În experiența mea, în domeniul științei este mai bine să nu-ți îndrepți toate eforturile spre a crea experimentul perfect. S-ar putea să complici lucrurile inutil. Reușitele științifice sunt mai degrabă rezultatul unei acumulări de succese aproximative și a creșterii treptate a încrederii.

Să aplicăm acum conștiinței aceeași abordare imperfectă, dar pragmatică.

Să presupunem că așezăm un subiect în fața unui

[0] M.S. Beauchamp, J.V. Haxby, J.E. Jennings și E.A. Deyoe, „An FMRI Version of the Farnsworth-Munsell 100-Hue Test Reveals Multiple Color – Selective Areas în Human Ventral Occipitotemporal Cortex”, Cerebral Cortex 9 (1999): 257 – 263.

[0] S.E. Bouvier și S.A. Engel, „Behavioral Deficits and Cortical Damage Loci în Cerebral Achromatopsia”, Cerebral Cortex 16 (2006): 183 – 191.

monitor. Acesta se uită fix la un punct din centrul ecranului. Dintr-odată, și doar pentru o clipă, vede imaginea unei fețe. Este vagă și nu e afișată decât pentru a douăzecea parte dintr-o secundă. După ce fața dispare, pătrate de culori aleatorii umplu ecranul. Această mască vizuală, după cum este numită, întrerupe procesarea cerebrală a feței și îngreunează văzul. Imaginea feței pâlpâie la marginea conștiinței. O dată la câteva secunde, o altă față este afișată în aceleași condiții dificile. Uneori subiectul spune că da, a văzut fața (să numim această circumstanță condiția A), și alteori spune că nu, nu a văzut o față (să-i spunem condiția B).

Avem aici un instrument simplu, imperfect, pentru studierea conștiinței în creier. În condiția A, participantul este conștient vizual de față și de masca în culori care acoperă ecranul după afișarea imaginii. În condiția B, e conștient doar de mască, nu de față. Într-un anumit sens, în condiția A are o conștiință vizuală mai amplă sau e conștient de imagini vizuale mai variate. Dacă găsim o regiune cerebrală mai activă în condiția A, putem presupune că această regiune cerebrală participă la conștiința vizuală. Ar fi oricum un prim indiciu. Experimentul ar putea indica alte regiuni cerebrale, care n-au legătură cu conștiința vizuală, dar măcar am avea un pas inițial în căutarea răspunsului și am putea trece apoi la alte experimente, cum ar fi studierea pacienților care au suferit accidente vasculare ce au afectat aceste regiuni cerebrale.

Se poate ca experimentul cu fețele să nu fi fost realizat niciodată exact așa cum l-am descris, dar se încadrează în categoria generală a experimentelor asupra conștiinței vizuale care au fost făcute de multe ori.[0] În condiția A,

[0] M. Binder, K. Gociewiez, B. Windey, M. Koculak, K. Fine, J. Nikadon, M. Derda și A. Cleeremans, „The Levels of Perceptual Processing and the Neural Correlates of Increasing Subiective Visibility”, Consciousness and Cognition 55 (2017): 106 – 125; D. Carmel, N. Lavie și G. Rees, „Conscious Awareness of Flicker în Humans Involves Frontal and Parietal Cortex”, Current Biology 16 (2006): 907 – 911; M.S. Christensen, T.Z. Rams Oy, T.E. Lund, K.H. Madsen și J.B. Rowe, „An

participantul este conștient de o imagine vizuală specifică. În condiția B e prezentată aceeași imagine sau o alta asemănătoare, dar participantul nu e conștient de ea.

Te-ai putea gândi că, în condiția A, când fața ajunge în conștiință, sistemul vizual procesează imaginea feței, în timp ce în condiția B, când fața nu ajunge în conștiință, e evident că sistemul vizual nu reacționează la ea. Dar nu asta se întâmplă de obicei. Regiunile cerebrale implicate în procesarea formei, culorii, a texturii și a altor proprietăți vizuale se activează când apare imaginea, indiferent dacă persoana o vede sau nu în mod conștient. Există totuși unele diferențe semnificative între condițiile de conștientizare și neconștientizare. Activitatea sistemului vizual tinde să fie susținută și mai puțin variabilă în cazul conștientizării decât în cazul neconștientizării.[0] Această diferență de activitate se potrivește ideii generale că, atunci când atenția amplifică și stabilizează semnalele vizuale, acestea au mai multe șanse să ajungă la conștiință. Cea mai mare diferență apare însă în anumite arii cerebrale din lobii parietal și frontal.[0] Aceste regiuni sunt activate

FMRI Study of the Neural Correlates of Graded Visual Perception", Neuroimage 31 (2006): 1711 – 1725; S. Dehaene și J.P. Changeux, „Experimental and Teoretical Approaches to Conscious Processing", Neuron 70 (2011): 200 – 227; S. Dehaene, L. Naccache, L. Cohen, D.L. Bihan, f. F. Mangin, J.B. Poline și D. Rivière, „Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming", Nature Neuroscience 4 (2001): 752 – 758.

[0] A. Schurger, I. Sarigiannidis, L. Naccache, J.D. Sitt și S. Dehaene, „Cortical Activity is More Stable When Senzory Stimuli are Consciously Perceived", Proceedings of the National Academy of Sciences U.S.A., 112 (2015): E2083 – 2092.

[0] M. Binder, K. Gociewiez, B. Windey, M. Koculak, K. Fine, J. Nikadon, M. Derda și A. Cleeremans, „The Levels of Perceptual Processing and the Neural Correlates of Increasing Subiective Visibility", Consciousness and Cognition 55 (2017): 106 – 125; D. Carmel, N. Lavie și G. Rees, „Conscious Awareness of Flicker în Humans Involves Frontal and Parietal Cortex", Current Biology 16 (2006): 907 – 911; M.S. Christensen, T.Z. Rams Oy, T.E. Lund, K.H. Madsen și J.B. Rowe, „An FMRI Study of the Neural Correlates of Graded Visual Perception", Neuroimage 31 (2006): 1711 – 1725; S. Dehaene și J.P. Changeux, „Experimental and Teoretical Approaches to Conscious Processing",

intens de imaginea vizuală în condiția A, când subiectul spune că e conștient de existența feței, dar sunt mult mai puțin activate, sau nu sunt activate deloc, în condiția B, când subiectul nu e conștient de existența feței.

Acest rezultat a fost repetat de multe ori cu multe variații, ceea ce mă face să am încredere în descoperiri. Unul dintre experimentele mele recente a dus la un rezultat aproape identic.[0] Când ți se arată ceva și devii conștient de asta, rețelele parietal-frontale reacționează. Când ți se arată ceva și sistemul tău vizual procesează imaginea respectivă, dar aceasta nu ajunge în conștiința ta, aceleași rețele tind să nu reacționeze. Cea mai simplă interpretare este că rețelele parietal-frontale sau un subsistem din cadrul lor pot construi conștiința.

O anumită parte a complexului parietal-frontal este indicată adesea ca posibil punct central al conștiinței. Cortexul prefrontal – partea cea mai din față a cortexului, aflată chiar înapoia frunții – este încă puțin cunoscut, dar ce se știe despre el arată că e un procesor extrem de sofisticat de informații. Principala calitate a cortexului prefrontal este flexibilitatea.[0] Se pare că e implicat în aproape orice tip de activitate. Dacă numeri puncte, neuronii din cortexul tău prefrontal țin șirul numerelor.[0] Dacă urmărești localizarea

---

Neuron 70 (2011): 200 – 227; S. Dehaene, L. Naccache, L. Cohen, D.L. Bihan, J.F. Mangin, J.B. Poline și D. Riviere, „Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming”, Nature Neuroscience 4 (2001): 752 – 758.

[0] T.W. Webb, K. Igelström, A. Schurger și M.S.A. Graziano, „Cortical Networks Involved în Visual Awareness Independently of Visual Attention”, Proceedings of the National Academy of Sciences U.S.A. 113 (2016): 13923 – 13928.

[0] T.J. Buschman și E.K. Miller, „Goal-Direction and Top-Down Control”, Philosophical Transactions of the Royal Society of London, B, Biological Sciences 369 (2014): 20130471; E.K. Miller și J.D. Cohen, „An Integrative Theory of Prefrontal Cortex Function”, Annual Review of Neurosciences 24 (2001): 167 – 202.

[0] A. Nieder și E.K. Miller, „Coding of cognitive magnitude: compressed scaling of numerical information în the primate prefrontal cortex”, Neuron 37 (2003): 149 – 57.

unor puncte pe ecranul unui computer, aceiași neuroni monitorizează poziția acestora.[0] într-un celebru experiment realizat de Earl Miller la MIT, niște maimuțe au fost învățate să catalogheze desene ambigue de animale ca pisici sau câini.[0] Neuronii din cortexul prefrontal au devenit rapid detectori de pisici și câini, unii reacționând când maimuța vedea o pisică, alții reacționând când vedea un câine. Sarcina era total arbitrară – maimuțele nu văzuseră niciodată în realitate o pisică sau un câine –, ceea ce arată că neuronii prefrontali pot prelua orice caracteristică, în funcție de sarcina care trebuie îndeplinită.

Specialiștii în neuroștiințe numesc uneori cortexul prefrontal, în special regiunea mare numită cortex prefrontal dorso-lateral, zona cerebrală responsabilă de „memoria de lucru”.[0] Este evident că informații din alte regiuni cerebrale pot fi încărcate în zona prefrontală, unde sunt menținute temporar și manipulate pentru a servi necesităților de moment.

Date fiind aceste proprietăți, cortexul prefrontal pare un candidat firesc la statutul de sediu al conștiinței.[0] Colectează informații prin conexiunile sale răspândite pe toată suprafața cortexului, ca un păianjen aflat în mijlocul pânzei.

Și totuși această soluție, oricât ar fi de tentantă, nu mă mulțumește. Sunt prea multe întrebări fără răspuns. De exemplu, când oamenii suferă leziuni la nivelul cortexului prefrontal, de ce nu-și pierd experiența conștiinței? Deși au

---

[0] S.C. Rao, G. Rainer și E.K. Miller, „Integration of What and Where în the Primate Prefrontal Cortex”, Science 276 (1997): 821 – 824.

[0] D.J. Freedman, M. Riesenhuber, T. Poggio și E.K. Miller, „Categorical Representation of Visual Stimuli în the Primate Prefrontal Cortex”, Science 291 (2001): 312 – 316.

[0] R. Levy R și P.S. Goldman-Rakic, „Segregation of Working Memory Functions Within the Dorso-lateral Prefrontal Cortex”, Experimental Brain Research 133 (2000): 23 – 32; E.K. Miller, „The «Working» of Working Memory”, Dialogues în clinical Neuroscience 15 (2013): 411 – 418.

[0] B. Ødegaard, R.T. Knight și H. L-au, „Should a Few Null Findings Falsify Prefrontal Theories of Conscious Perception?”, Journal of Neuroscience 40 (2017): 9593 – 9602.

dificultăți în a face planuri și a trece de la o sarcină la alta, în general nu-și pierd senzația că au conștiință.[0] Cel puțin nu în mod tipic.

Dar am o întrebare și mai simplă despre ipoteza cortexului prefrontal: când informația ajunge în această regiune cerebrală, de ce ar genera o percepție conștientă? Atâta timp cât neuroștiințele caută un sediu al conștiinței – o regiune centrală a creierului în care informațiile converg și generează conștiință –, cred că știința bate pasul pe loc.

E aceeași problemă pe care am subliniat-o mai devreme. Unde se duce informația despre formă pentru a genera culoarea? Nicăieri. Pentru a înțelege culoarea, trebuie să căutăm locul în care culoarea este procesată. Unde se duce informația vizuală pentru a genera o experiență conștientă? Nicăieri. Pentru a înțelege conștiința, trebuie să căutăm un sistem cerebral care procesează *informația* despre *conștiință* – despre proprietățile și consecințele ei. Informația despre conștiință, procesată într-un anumit sistem, este legată de informația vizuală despre obiectul la care te uiți, procesată într-un alt sistem. Cele două pachete de informații, unite într-o alianță temporară care se întinde pe mai multe rețele cerebrale, formează un model intern mai amplu, care spune, drept urmare: „Aici este experiența mea conștientă, aici este un măr și în acest moment cele două sunt legate între ele”.

Accept bucuros că o parte a acestui proces este cortexul prefrontal. Poate că funcționează ca forum mintal, așa cum presupun oamenii de știință, ca acumulare activă de obiective, gânduri și observații, contribuind la organizarea comportamentului nostru. Această idee este convingătoare și este susținută de date empirice. Dar eu susțin că acumularea semnalelor din creier trebuie să conțină pe undeva informații despre conștiință, altfel nu am putea niciodată spune ceva despre această proprietate. Știm deja ce rețele cerebrale procesează informația vizuală despre un măr. Sistemul vizual a fost studiat intens timp de aproape

[0] Mai multe analize ample susțin că pierderea sau reducerea conștiinței nu este considerată în general un simptom al leziunilor prefrontale.

un secol. Dar unde sunt rețelele care procesează informația despre conștiință?

Colegii mei din domeniul neuroștiințelor știu cât sunt de vag când vorbesc despre rețelele care se întind de la lobul parietal la lobul frontal. Cel puțin 12 au fost studiate în detaliu.[0] Printre acestea, se numără rețeaua dorsală a atenției, rețeaua ventrală a atenției, rețeaua de evidențiere

[0] Cercetările privind rețelele parietal-frontale s-au înmulțit foarte mult în ultima vreme. Menționez în continuare un minimum de surse despre principalele rețele discutate în text. D. Bzdok, R. Langner, L. Schilbach, O. Jakobs, C. Roski, S. Caspers, A.R. Laird, P.T. Fox.
K. Zilles și S.B. Eickhoff, „Characterization of the Temporo-Parietal Junction By Combining Data-Driven Parcellation, Complementary Connectivity Analyses, and Functional Decoding”, Neuroimage 81 (2013): 381 – 392; M. Corbetta, G. Patel și G.L. Shulman, „The Reorienting Sistem of the Human Brain: From Environment to Theory of Mind”, Neuron 58 (2008): 306 – 324; N.U. Dosenbach, D.A. Fair, F.M. Miezin, A.L. Cohen, K.K. Wenger, R.A. Dosenbach, M.D. Fox, A.Z. Snyder, J.L. Vincent, M.E. Raichle, B.L. Schlaggar și S.E. Petersen, „Distinct Brain Networks for Adaptive and Stable Task Control în Humans”, Proceedings of the National Academy of Sciences U.S.A. 104 (2007): 11073 – 11078; M.D. Fox, M. Corbetta, A.Z. Snyder, J.L. Vincent și M.E. Raichle, „Spontaneous Neuronal Activity Distinguishes Human Dorsal and Ventral Attention Sistems”, Proceedings of the National Academy of Sciences U.S.A. 103 (2006): 10046 – 10051; K. Igelström și M.S.A. Graziano, „The Inferior Parietal Lobe and Temporoparietal Junction: A Network Perspective”, Neuropsychologia 105 (2017): 70 – 83; K.M. Igelström, T.W. Webb și M.S.A. Graziano, „Neural Processes în the Human Temporoparietal Cortex Separated by Localizând Independent Component Analysis”, Journal of Neuroscience 35 (2015): 9432 – 9445; K.M. Igelström, T.W. Webb și M.S.A. Graziano, „Topographical Organization of Attentional, Social and Memory Processes în the Human Temporoparietal Cortex”, eneuro 3 (2016): e0060; R.B. Mars, f. Sallet, U, Schüffelgen, S. Fbabdi, I. Toni și M.F. Rushworth, „Connectivity-Based Subdivisions of the Human Right Temporoparietal Juncționarea: Evidence for Different Areas Participating în Different Cortical Networks”, Cerebral Cortex 22 (2012): 1894 – 1903; R. Ptak, „The Frontoparietal Attention Network of the Human Brain: Action, Saliency, and a Priority Pap of the Environment”, Neuroscientist 18 (2012): 502 – 515; R. Saxe și L.J. Powell, „It’s the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind”, Psychological Science 17 (2006): 692 – 699; J.L.

(*salience network*), rețeaua de control, rețeaua teoriei minții și rețeaua modului standard. Și mai sunt și altele, în unele cazuri cu funcții extrem de specifice. O rețea are legătură cu controlul mișcărilor oculare și se suprapune probabil controlului atenției.[0] O alta controlează direcția[0]. Altă rețea controlează forma mâinilor atunci când apucăm ceva.[0] în laboratorul meu, înainte să mă dedic cercetării conștiinței, am studiat rețeaua parietal-frontală care

---

Vincent, I. Kahn, A.Z. Snyder, M.E. Raichle și R.L. Buckner, „Evidence For a Frontoparietal Control Sistem Revealed by Intrinsic Functional Connectivity”, Journal of Neurophysiology 100 (2008): 3328 – 3342; M. Școlari, K.N. Seidl-Rathkopf și S. Kastner, „Functions of the Human Frontoparietal Attention Network: Evidence From Neuroimaging”, Current Opinion în Behavioral Sciences 1 (2015): 32 – 39; B.T.T. Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, D. Lashkari, M. Hollinshead, J.L. Roffman, J.W. Smoller, L. Zollei, J.R. Polimeni, B. Fischl, H. Liu și R.L. Buckner, „The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity”, Journal of Neurophysiology 106 (2011): 1125 – 1165.

[0] C. Amiez și M. Petrides, „Anatomical Organization of the Eye Fields în the Human and Non-Human Primate Frontal Cortex”, Progress în Neurobiology 89 (2009): 220 – 230; L.L. Chen și E.J. Tehovnik, „Cortical Control of Eye and Head Movements: Integration of Movements and Percepts”, European Journal of Neuroscience 25 (2007): 1253 – 1264; M.H. Grosbras și A. Berthoz, „Parieto-Frontal Networks and Gaze Shifts în Humans: Review of Funcțional Magnetic Resonance Imaging Data”, Advances în Neurology 93 (2003): 269 – 280; E. Lobei, P. Kahane, U. Leonards, M. Grosbras, S. Lehericy, D. Le Bihan și A. Berthoz, „Localization of Human Frontal Eye Fields: Anatomical and Functional Findings of Funcțional Magnetic Resonance Imaging and Intracerebral Electrical Stimulation”, Journal of Neurosurgery 95 (2001): 804 – 815.

[0] R. Caminiți, S. Ferraina și P.B. Johnson, „The Sources of Visual Information to the Primate Frontal Lobe: A Novel Role For the Superior Parietal Lobule”, Cerebral Cortex 6 (1996): 319 – 328; C.S. Konen, R.E. Mruczek, J.L. Montoya și S. Kastner, „Funcțional Organization of Human Posterior Parietal Cortex: Graspingând Reaching-Related Activations Relative to Topographically Organized Cortex”, Journal of Neurophysiology 109 (2013): 2897 – 9208; L.H. Snyder, A.P. Batista și R.A. Andersen, „Coding of Intention în the Posterior Parietal Cortex”, Nature 386 (1997): 167 – 170.

[0] M.G. Di Bono, C. Begliomini, U. Castiello și M. Zorzi, „Probing the Reaching-Grasping Network în Humans Through Multivoxel Pattern Decoding”, Brain and Behavior 5 (2015): e00412; C.S. Konen, R.E.

procesează zona spațială din apropierea corpului, spațiul peripersonal, și contribuie la coordonarea acțiunilor defensive în raport cu obiectele care se apropie.[0] O altă rețea are un rol în numărare și în raționamentele matematice.[0] Diversitatea este incredibilă. Pentru ca lucrurile să fie și mai complicate, multe dintre aceste rețele au granițe neclare și sunt recrutate de sarcini diverse, care se suprapun. Unde în acest haos de rețele și regiuni poate fi procesată conștiința?

Cea mai bună ipoteză pe care o am în prezent este că procesarea care duce la crearea constructului de conștiință este amplificată în special în zona cortexului numită joncțiunea temporo-parietală (JTP).[0] Creierul are două astfel

---

Mruczek, J.L. Montoya și S. Kastner, „Funcțional Organization of Human Posterior Parietal Cortex: Graspingând Reaching-Related Activations Relative to Topographically Organized Cortex”, Journal of Neurophysiology 109 (2013): 2897 – 9208; A. Murata, V. Gallese, G. Luppino, M. Kaseda și H. Sakata, „Selectivity for the Shape, Size, and Orientation of Objects for Grasping în Neurons of Monkey Parietal Area AI P”, Journal of Neurophysiology 83 (2000): 2580 – 2601; G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino și M. Matelli, „Funcțional Organization of Inferior Area 6 În The Macaque Monkey. II. Area F_5 and the Control of Distal Movements”, Experimental Brain Research 71 (1988): 491 – 507.

[0] D.F. Cooke și M.S.A. Graziano, „Super-Flinchers and Nerves of Steel: Defensive Movements Altered by Chemical Manipulation of a Cortical Motor Area”, Neuron 43 (2004): 585 – 593; D.F. Cooke, C.S.R. Taylor, T. Moore și M.S.A. Graziano, „Complex Movements Evoked by Microstimulation of Area VIP”, Proceedings of the National Academy of Sciences U.S.A. 100 (2003): 6163 – 6168.

[0] E. Eger, P. Pinel, S. Dehaene și A. Kleinschmidt, „Spatially Invariant Coding of Numerical Information în Functionally Defined Subregions of Human Parietal Cortex”, Cerebral Cortex 25 (2015): 1319 – 1329;
A. Nieder și E.K. Miller, „Coding of Cognitive Magnitude: Compressed Scaling of Numerical Information în the Primate Prefrontal Cortex”, Neuron 37 (2003): 149 – 157; R. Stănescu-Cosson, P. Pinel, P.F. Van de Moortele, D. Le Bihan, L. Cohen și S. Dehaene, „Understanding Dissociations în Dyscalculia: A Brain Imaging Study of the Impact of Number Size on the Cerebral Networks for Exact and Approximate Calculation”, Brain 123 (2000): 2240 – 2255.

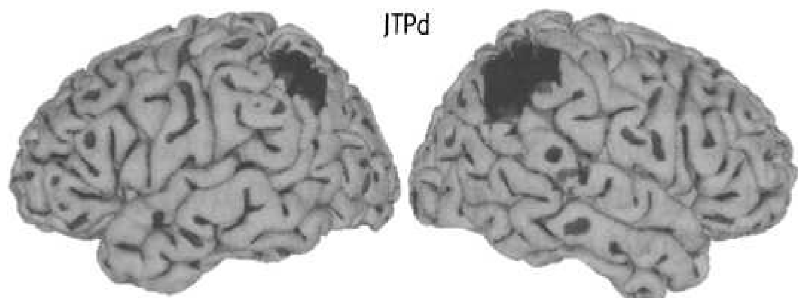
[0] K.M. Igelström, T.W. Webb și M.S.A. Graziano, „Topographical Organization of Attentional, Social and Memory Processes în the

de joncțiuni, una de fiecare parte, mai mult sau mai puțin deasupra urechilor și pe suprafața cortexului. Joncțiunea de pe partea dreaptă a creierului poate fi mai mare sau mai bine dezvoltată decât cea de pe stânga, deși cred că natura exactă a asimetrici nu e clară. În fiecare emisferă cerebrală, JTP este ea însăși împărțită în subregiuni cu proprietăți diferite, și cea mai de interes pentru mine este diviziunea situată cel mai sus, cea care se află în mai mare parte în lobul parietal și mai puțin în cel temporal. În laboratorul meu, numim această regiune JTP dorsală sau JTPD (vezi Figura 6.1). Este uneori numită și lobul parietal inferior. JTPD se află la intersecția dintre rețeaua de control, rețeaua ventrală a atenției, rețeaua de evidențiere și rețeaua teoriei minții.[0] Este unul dintre centrele cerebrale cel mai amplu conectate la informație, sub acest aspect rivalizând chiar cu cortexul prefrontal. Dacă această regiune contribuie la procesarea conștiinței, pesemne că nu o face singură, dat fiind că este probabil ca procesările complexe să depindă de rețele răspândite, nu de regiuni izolate. E, de asemenea, probabil să aibă multe alte funcții în afară de procesarea conștiinței. În cele ce urmează mă voi referi la JTP în general, și nu la subregiunea JTPD, pentru că nu avem suficiente informații pentru a exclude celelalte componente

Human Temporoparietal Cortex", eneuro 3 (2016): e0060; Y.T. Kelly, T.W. Webb, J.D. Meier, MJ. Arcaro și M.S.A. Graziano, „Attributing Awareness to Oneself and to Others", Proceedings of the National Academy of Sciences U.S.A. 111 (2014): 5012 – 5017; T.W. Webb, K. Igelström, A. Schurger și M.S.A. Graziano, „Cortical Networks Involved în Visual Awareness Independently of Visual Attention", Proceedings of the National Academy of Sciences U.S.A. 113 (2016): 13923 – 13928.

[0] K. Igelström și M.S.A. Graziano, „The Inferior Parietal Lobe and Temporoparietal Junction: A Network Perspective", Neuropsychologia 105 (2017): 70 – 83; K.M. Igelström, T.W. Webb și M.S.A. Graziano, „Topographical Organization of Attentional, Social and Memory Processes în the Human Temporoparietal Cortex", eneuro 3 (2016): e0060; R.B. Mars, J. Sallet, U. Schuffelgen, S. Jbabdi, I. Toni și M.F.S. Rushworth, „Connectivity-Based Subdivisions of the Human Right Temporoparietal Juncționarea: Evidence for Different Areas Participating în Different Cortical Networks", Cerebral Cortex 22 (2012): 1894 – 1903.

ale JTP. Presupunem însă că JTPD poate juca un rol mai mare sau central în procesarea conștiinței.



6.1 Segmentul superior sau dorsal al joncțiunii temporo-parietale (JTPD) în partea stângă și în partea dreaptă a cortexului cerebral uman, identificat prin convergența a doua metode: prima, o tehnică numită analiză independentă a componentelor, și a doua, tiparul de conectivitate cu restul cortexului. (Retipărit de la p. 9435 din K.M. Igelström, T.W. Webb și M.S.A. Graziano, „Neural Processes în the Human Temporoparietal Cortex Separated By Localizând Independent Component Analysis”, în *Journal of Neuroscience*, nr. 35, pp. 9432 – 45, 2015.)

Vreau să fie clar ceea ce propun. Eu nu susțin că JTP ar fi în sine conștientă de ceva anume și nici că ar genera experiența conștientă. Nu e un homuncul – un omuleț pitit în țeastă. Este o parte a unei rețele care alcătuiește un construct, un model care-i spune creierului ce este conștiința. Fără această informație nu am fi capabili să ne atribuim nouă înșine conștiință, nu am putea să susținem că avem conștiință și, dacă am fi întrebați despre asta, nici măcar nu am înțelege întrebarea.

Am mai multe motive pentru a pune accentul pe această regiune. În primul rând, experimentele din ultimii 30 de ani au arătat că este implicată în procesarea stărilor mintale posibile ale celorlalți oameni.[0] Când te gândești la credințele, emoțiile sau intențiile altora, rețeaua ta JTP se

[0] R. Saxe și L.J. Powell, „It’s the Thought that Counts: Specific Brain Regions for One Component of Theory of Mind”, Psychological Science 17 (2006): 692 – 699.

activează ca nod central în rețeaua teoriei minții. Dacă încercăm să găsim regiunile cerebrale care ne pot oferi însăși noțiunea de minte conștientă, JTP este locul firesc în care să căutăm. În experimentele mele de laborator, când oamenii se gândesc la conștiință, indiferent dacă răspund la întrebări despre propria conștiință („Da, sunt conștient de acea imagine”) sau dacă le atribuie altora stări de conștiință („Da, cred că Kevin este conștient de acel obiect de lângă el”), JTP se activează ca și cum ar participa la constructul de conștiință[0].

Un al doilea motiv pentru care presupun că JTP poate fi implicată în conștiință este relația pe care o are cu atenția. Să ne amintim că, potrivit teoriei schemei atenției, creierul construiește informații despre conștiință dintr-un motiv anume. Aceste informații funcționează ca schemă utilă a atenției, un model care descrie unele dintre proprietățile atenției. Orice rețea corticală care contribuie la controlul atenției trebuie să aibă acces la o schemă a atenției. Așadar, ar trebui să căutăm rețelele care au legătură cu atenția – rețeaua dorsală a atenției, rețeaua ventrală a atenției, rețeaua de evidențiere, rețeaua de control. Aceste rețele sunt considerate a constitui motorul atenției, fiind capabile să influențeze alte regiuni ale cortexului pentru a amplifica unele semnale în defavoarea altora.[0] Înspre acestea se îndreaptă în mod firesc presupunerile noastre legate de construirea unei scheme a atenției. Într-o serie de

---

[0] Y.T. Kelly, T.W. Webb, J.D. Meier, M.J. Arcaro și M.S.A. Graziano, „Attributing Awareness to Oneself and to Others”, Proceedings of the National Academy of Sciences U.S.A. 111 (2014): 5012 – 5017; T.W. Webb. K. Igelström, A. Schurger și M.S.A. Graziano, „Cortical Networks Involved în Visual Awareness Independently of Visual Attention”, Proceedings of the National Academy of Sciences U.S.A. 113 (2016): 13923 – 13928.

[0] M. Corbetta, G. Patel și G.L. Shulman, „The Reorienting Sistem of the Human Brain: From Environment to Theory of Mind”, Neuron 58 (2008): 306 – 324; T. Moore și M. Zimsak, „Neural Mechanisms of Selective Visual Attention”, Annual Review în Psychology 68 (2017): 47 – 72. R. Ptak, „The Frontoparietal Attention Network of the Human Brain: Action, Saliency, and a Priority Map of the Environment”, Neuroscientist, 18 (2012): 502 – 515.

experimente, inclusiv în cele desfășurate în laboratorul meu, activitatea cerebrală asociată cu atenția și activitatea cerebrală asociată cu cognițiile sociale prezintă o suprapunere într-o zonă anume a JTP, mai ales în jumătatea superioară a acesteia.[0]

O rezervă pe care am auzit-o în legătură cu implicarea JTP în conștiință este că, atunci când devii conștient de o imagine vizuală, cum ar fi imaginea unei fețe, nu pare să existe nicio dovadă că această informație ar ajunge la JTP – sau măcar la vreuna dintre regiunile lobului parietal. Identitatea unei fețe, expresia afectivă a feței, micile detalii vizuale care alcătuiesc fața – niciuna dintre aceste informații nu pare să fie procesată în lobul parietal. Punând bazele unei teorii devenite clasice, în anii 1990 psihologii David Milner și Mel Goodale[0] au susținut că cortexul parietal *nu* este conștient vizual, în timp ce alte regiuni ale sistemului vizual, în special lobul temporal, *sunt* conștiente vizual.

Alți oameni de știință nu sunt de acord nici cu lobul parietal, nici cu cel temporal ca surse ale conștiinței vizuale, afirmând că, de fapt, conștiința este conectată la nivelurile inferioare de procesare, unde informația vizuală pătrunde în cortex prin regiunea V1.[0] În acest moment, când mă uit la ecranul computerului, sunt conștient de curbele și unghiurile literelor pe care le scriu. Dacă mă uit dincolo de ecran, văd pe fereastră nenumărate frunze de sicomor. Culorile sunt intense și ies în evidență texturile și cele mai mici detalii. Regiunea V1 extrage tocmai acest tip

---

[0] K.M. Igelström, T.W. Webb și M.S.A. Graziano, „Topographical Organization of Attentional, Social and Memory Processes în the Human Temporoparietal Cortex”, eneuro 3 (2016): e0060.

[0] M.A. Goodale și A.D. Milner, „Separate Visual Pathways for Perception and Action”, Trends în Neurosciences 15 (1992): 20 – 25.

[0] M. Hurme, M. Koivisto, A. Revonsuo și H. Railo, „Early Processing în Primary Visual Cortex is Necessary for Conscious and Unconscious Vision While Late Processing is Necessary Only for Conscious Vision în Neurologically Healthy Humans”, Neuroimage 150 (2017): 230 – 238; F. Tong, „Primary Visual Cortex and Visual Awareness”, Nature Reviews Neuroscience 4 (2003): 219 – 29.

de detalii. Mai mult, când oamenii suferă leziuni la nivelul V1, își pierd percepția vizuală conștientă.[0] Date fiind aceste observații, conștiința se poate să nu fie generată în partea superioară a ierarhiei corticale, unde se consideră că au loc cogniția și procesele superioare de gândire, ci s-ar putea să fie la bază, unde sunt procesate detaliile brute.

Perspectivele expuse anterior par să contrazică propunerea mea că JTP ar fi implicată în conștiință. Însă teoria schemei atenției restructurează presupunerile de la care pornim și face toate aceste perspective compatibile unele cu celelalte. Să presupunem că ești conștient vizual de imaginea unei fețe. Zona V1 contribuie la construirea informației despre detalii și culoare, saturație și intensitate. Lobul temporal construiește informația despre identitatea feței. JTP contribuie la construirea informației despre conștiință. E nevoie de toate aceste elemente ca să poți susține că ești conștient de o față. Milner și Goodale au avut dreptate când au spus că *conținutul* conștiinței vizuale, în acest caz detaliile și identitatea feței, nu se găsesc de obicei în regiunile parietale. Dar fără JTP s-ar putea să nu existe un construct de conștiință care să fie atașat acestor conținuturi.

O modalitate tradițională de conceptualizare a sistemului vizual constă în comparația cu o linie de producție. Pachete de informații intră prin ochi și trec apoi de la un stadiu de procesare la următorul, urcând în ierarhie, din ce în ce mai elaborate. Odată ce produsul ajunge la nivelul cel mai înalt, procesul de construire se va fi încheiat, iar stadiile anterioare, dat fiind că și-au îndeplinit diferitele sarcini, sunt acum scoase din uz. Nivelul cel mai înalt al ierarhiei corticale poate să ia această informație pe deplin asamblată și să o transforme în acțiuni sau în limbaj – sau în conștiință.

O analogie mai potrivită ar fi comparația cu teancul de

---

[0] A. Cowey, „The Blindsight Saga”, Experimental Brain Research 200 (2010): 3 – 24; L. Weiskrantz, E.K. Warrington, M.D. Sanders și f. Marshall, „Visual Capacity în the Hemianopic Field Following a Restricted Cortical Ablation”, Brain 97 (1974): 709 – 728.

schițe ale unei case. Fiecare pagină descrie clădirea într-un mod diferit, la un alt nivel de detaliu. O pagină indică amplasarea pereților. O alta indică sistemul electric. Altă pagină descrie instalația sanitară și o alta dulapurile montate pe pereți. Există chiar și o schiță amănunțită a mobilierului recomandat. Schițele formează o ierarhie naturală, dat fiind că nu poți să întorci pagina fără să fi înțeles detaliile din paginile anterioare, într-un sens, informația curge de la o pagină la următoarea, din ce în ce mai cizelată. Dar, în același timp, pe măsură ce înaintezi, nu arunci paginile deja parcurse pentru un plan ultim, o unică pagină care cuprinde toate informațiile. Fiecare conține de fapt ceva util și dacă pierzi vreuna nu poți construi casa. Acțiunea finală, construirea casei, depinde de fiecare pagină și de toate la un loc, nu doar de pagina ultimă, care cuprinde nivelul maxim de procesare a informației.

La fel, cortexul vizual are o ierarhie – probabil mai multe ierarhii care interacționează.[0] Dar niciunul din stadiile superioare nu înlocuiește un stadiu anterior. Nu pare să existe un stadiu ultim în care toate informațiile importante ajung la capătul călătoriei pentru a intra în conștiință și pentru a declanșa acțiunea. În schimb, fiecare stadiu ierarhic oferă o analiză diferită, renunțând la unele tipuri de informație, îmbogățind altele, adăugând straturi peste straturi de analiză utilă a lumii vizuale. Această acumulare de informație, de la micile detalii vizuale de la nivelurile inferioare la informația conceptuală profundă de la nivelurile superioare, influențează acțiunea și vorbirea. Outputul nu este determinat doar de cel mai înalt nivel al

---

[0] T.N. Aflalo și M.S.A. Graziano, „Organization of the Macaque Extrastriate Visual Cortex Re-examined Using the Principle of Spațial Continuity of Function”, Journal of Neurophysiology 105 (2011): 305 – 320; D. Felleman și D. Van Essen, „Distributed Hierarchical Processing în the Primate Visual Cortex”, Cerebral Cortex 1 (1991): 1 – 47; M.A. Goodale și A.D. Milner, „Separate Visual Pathways for Perception and Action”, Trends în Neurosciences 15 (1992): 20 – 25; L.G. Ungerleider și J.V. Haxby, „ «What» and «Where» în the Human Brain”, Current Opinion în Neurobiology 4 (1994): 157 – 165.

ierarhiei, ci de toate nivelurile simultan.

Dacă cineva spune: „În fotografie e fața prietenului meu”, aserțiunea se bazează pe informația analizată în regiunile care se ocupă de procesarea feței în cele mai înalte niveluri ale ierarhiei vizuale, situate în lobul temporal. Dacă spune: „Prietenul meu are o pată pe nas”, acest comentariu depinde probabil de informația din regiunile vizuale de nivel inferior, unde sunt procesate mai degrabă detaliile vizuale. Dacă spune: „Prietenul meu are părul șaten”, acest comentariu se bazează pe informația din regiunile de procesare a culorii din centrele specializate în culoare din cortexul vizual. Dacă spune: „Prietenul meu are vocea ascuțită”, acest comentariu se poate baza pe informația obținută din memorie și reluată într-o parte a cortexului auditiv. Dacă spune: „Mâine după-amiază îi voi da înapoi cei 20 de dolari pe care i-i datorez”, observația poate depinde de informația construită parțial în cortexul prefrontal, care se ocupă de planificare.

Iar dacă cineva spune: „Am o *experiență subiectivă*. Am un misterios sine subiectiv, o conștiință”, această afirmare a conștiinței se poate să depindă de informația creată într-o altă rețea corticală specializată, care se poate intersecta cu JTP.

Nu vreau să spun că cortexul este o serie de module separate, fiecare limitat la un anumit domeniu de informație și izolat de celelalte. De fapt, deși sunt specializate în diferite tipuri de informație, regiunile corticale interacționează și se influențează reciproc într-o rețea masivă, rezonantă. Atingi o coardă a rețelei – adaugi o nouă informație vizuală în V1 s-au adaugi un nou gând construit în cortexul prefrontal – și reverberează în întreaga rețea, provocând tipare de activitate aflate în permanentă schimbare. Pachetele de informații sunt amplificate de atenție sau se întorc și se contopesc în zgomotul din fundal. Regiunile corticale formează constant noi alianțe, deschizând linii temporare de comunicare, combinându-și informațiile în structuri mai mari, apoi decuplându-se pentru a forma alte alianțe noi.

În acest cadru de referință nu trebuie să stabilim dacă conștiința este asociată mai degrabă cu nivelurile de la baza ierarhiei, cu nivelurile superioare sau dacă e undeva la mijloc. Putem fi conștienți de detalii vizuale procesate în V1 și la fel de conștienți de abstracțiuni procesate în cortexul prefrontal. Motivul este că noțiunea de conștiință e definită de o serie de informații construite într-o rețea specializată și că această rețea poate crea o alianță funcțională cu diferite regiuni corticale.

Să ne imaginăm un extraterestru inteligent care vizitează Pământul. Să presupunem că creierul său nu are un construct al conștiinței care să aducă laolaltă informațiile interne pe care le deține. Îi lipsește echivalentul zonei JTP a oamenilor. Totuși, probabil că are modele ale sinelui – poate să-și monitorizeze propriile procese interne și să-și controleze comportamentul –, dar a evoluat diferit și nu deține noțiunea de conștiință subiectivă.

Îl întrebăm:

— Vezi mărul?

Extraterestrul răspunde:

— Da, îl văd.

— Ai o experiență conștientă a mărului?

— Îl procesez vizual.

— Dar mai ai și altceva, acel *je ne sais quoi*, o esență a experienței asociate mărului, sau doar procesezi informația? *Simți* cumva procesarea mărului?

— Nu știu ce este acel *je ne sais quoi*. Mărul există, și eu îl procesez vizual. Ce altceva să fac? De ce ar exista o esență? Din ce e făcută esența și de unde vine?

— Aha! Dar folosești cuvântul „eu”, deci trebuie să ai un concept de sine.

— Da, așa e. Conversația asta se desfășoară între doi agenți care pot fi numiți „tu” și „eu”. Am o serie de informații relativ completă despre corpul meu fizic, despre trecutul meu și despre caracteristicile mele comportamentale.

— Deci până la urmă ai conștiință de sine?

— Am informații despre mine. Ce e conștiința?

— E atunci când, pe lângă procesarea informației despre tine, ai o experiență subiectivă a acesteia. Mai ai ceva, un *je ne sais quoi*, o… la naiba, nu contează.

Vreau să spun că mintea este informație. Îmi place să spun că mintea umană este o sculptură din trilioane de straturi de informații, aflată permanent în schimbare, minunat de complicată. Dacă mintea nu are informații despre conștiință, nu poate ști ce este această proprietate și nu și-o poate atribui și nu o poate atribui altora. Noi, oamenii, facem caz de conștiință doar pentru că avem un subsistem cerebral care construiește informații despre ea. Iar eu cred că știința se apropie de localizarea acestui subsistem și de identificarea modului în care e conectat la alte sisteme.

O modalitate tradițională de localizare a funcțiilor creierului este studierea consecințelor leziunilor cerebrale. Cortexul vizual, de exemplu, a fost identificat când, la maimuțe, vătămări ale părții din spate a creierului, în regiunea cunoscută acum ca V1, au provocat orbirea.[0] Centrele vorbirii din creierul uman au fost descoperite când leziunile de la nivelul unei părți a lobului frontal, cunoscută acum ca aria Broca sau aria vorbirii, au anulat capacitatea oamenilor de a vorbi.[0] Ariile cerebrale care identifică obiectele în mișcare au fost descoperite când leziunile de la nivelul unei anumite părți a cortexului vizual au dus la

---

[0] S. Brown și E. Schäfer, „An Investigation into the Functions of the Occipital and Temporal Lobes of the Monkey’s Brain”, Philosophical Transactions of the Royal Society of London, B, Biological Sciences 179 (1888): 303 – 327.

[0] P. Broca, „Remarks on the Seat of the Faculty of Articulate Language, Followed by an Observation of Aphemia”, Bulletin de la Société Anatomique de Paris 6 (1861): 330 – 357, trad. De G. Von Bonin, republicată în G. Von Bonin (coord.), Some Papers On The Cerebral Cortex (Springfield, IL: Charles Thomas Publisher, 1960, Pages 49 – 72); A.R. Damasio și N. Geschwind, „The Neural Basis of Language”, Annual Review of Neuroscience? (1984): 127 – 147.

incapacitatea de a vedea lucrurile aflate în mișcare.[0] O pacientă celebră cu această afecțiune nu putea să toarne apă. Nu vedea decât un țurțure care atârna de vas și apoi o baltă pe masă. Iar leziunile de la nivelul regiunilor corticale responsabile de procesarea culorii duc la incapacitatea de-a percepe culorile.[0]

Există leziuni ale cortexului care să elimine conștiința?

Nu căutăm o afecțiune care îi face pe oameni să adoarmă. Ciclul de somn și de veghe este controlat de structuri mult mai vechi, situate în trunchiul cerebral.[0] Ce căutăm este o afecțiune în care oamenii sunt practic treji, informația poate pătrunde în sistemul cortical și le poate chiar influența comportamentul, dar nu apare nicio conștiință subiectivă a acestei informații.

Ceea ce căutăm nu este nici acel zombi imaginat de filosofi.[0] Filosofii au inventat un concept mai degrabă specializat de zombi – nu chestia fioroasă care se bălăngăne cu mâinile întinse înainte, ci o persoană care arată și se poartă normal, chiar vorbește normal, deși nu are niciun fel de experiență subiectivă. Oamenii de știință încă discută dacă zombi există sau dacă ar putea să existe. Ipoteticul extraterestru pe care l-am descris mai sus este un astfel de zombi. Teoria schemei atenției admite posibilitatea existenței zombilor: poate că un zombi poate fi construit artificial sau poate a evoluat pe o altă planetă. Dar dacă teoria este corectă nu am putea niciodată să transformăm o persoană normală într-un astfel de zombi al

---

[0] J. Zihl, D. Von Cramon, N. Mai, „Selective Disturbance of Movement Vision After Bilateral Brain Damage”, Brain 106 (1983): 313 – 340.

[0] S.E. Bouvier și S.A. Engel, „Behavioral Deficits and Cortical Damage Loci în Cerebral Achromatopsia”, Cerebral Cortex 16 (2006): 183 – 191.

[0] C. Gottesmann, „The Neurophysiology of Sleep and Waking: Intracerebral Connections, Functioning and Ascending Influences of the Medulla Oblongata”, Progress în Neurobiology 59 (1999): 1 – 54.

[0] D. Chalmers, The Conscious Mind (Oxford, UK: Oxford University Press); P. Skokowski, „I, Zombie”, Consciousness and Cognition 11 (2002): 1 – 9; C. Tandy, „Are You (Almost) a Zombie? Conscious Thoughts About «Consciousness în the Universe» by Hameroff and Penrose”, Physics of Life Reviews 11 (2014): 89 – 90.

filosofilor, care să meargă și să vorbească normal, dacă i-am elimina o parte a creierului. Nu am putea să eliminăm conștiința fără să facem acea persoană incapabilă să funcționeze în viața de zi cu zi. Funcționarea normală a creierului uman depinde prea mult de constructul său de conștiință.

Dacă schema atenției ți-ar fi îndepărtată chirurgical, ai suferi de cel puțin trei afecțiuni. În primul rând, nu ai fi capabil să-ți reglezi propria atenție. Dacă nu ți-ai putea controla atenția, atunci ai fi incapacitat în mod fundamental, incapabil să-ți orientezi acțiunile într-un mod susținut spre atingerea unor obiective sau scopuri specifice. Ai fi chiar mai afectat decât zombi din filme, care, în ciuda mersului lor teleghidat, își urmăresc cu succes victimele și le devorează creierul. În al doilea rând, nu ai avea capacitatea de a construi modele sociale ale minții celorlalți. Fără capacitatea de a le atribui celorlalți conștiință, cogniția ta socială s-ar prăbuși. În al treilea rând, fără informațiile necesare, nu ai putea să înțelegi întrebările despre conștiință sau să faci afirmații coerente despre aceasta. Un zombi tradițional, precum cel al filosofilor, s-ar putea amesteca în mulțime, dar un zombi fără schema atenției ar fi un individ grav afectat. S-ar putea să reacționeze dacă îl înțepi cu un ac, dar cam atât.

În lumea clinică a leziunilor cerebrale există o afecțiune deosebit de gravă, numită sindrom de neglijare hemispațială, care a fost studiată timp de aproape 100 de ani.[0] Leziunile de la nivelul unei anumite părți a creierului duc la pierderea conștiinței obiectelor și a evenimentelor din partea opusă a corpului. Mecanismul afectat în cadrul acestei boli este prezent în ambele emisfere cerebrale, dar leziunile părții drepte duc la o mai serioasă și permanentă

[0] W.R. Brain, „A Form of Visual Disonentation Rezulting from Lesions of the Right Cerebral Hemisphere”, Proceedings of the Royal Society of Medicine 34 (1941): 771 – 776; M. Critchley, The Parietal Lobes (London: Hafner Press, 1953); G. Vallar, „Extrapersonal Visual Unilateral Spațial Neglect and its Neuroanatomy”, Neuroimage 14 (2001): S52-S58.

pierdere a funcțiilor.[0]

Neglijarea nu e orbire. În cazul orbirii, obiectele dispar din câmpul vizual, dar știi în continuare că există. Nimeni nu poate vedea ce are la spate, de exemplu, dar avem totuși o idee despre ce se află acolo. În cazul sindromului neglijării, obiectele pur și simplu dispar din conștiință când ajung în partea nepotrivită din spațiu. O atingere pe partea stângă a corpului nu este înregistrată. Un sunet care vine din stânga ori e atribuit în mod greșit părții drepte, ori e cu totul ignorat. În cazurile cele mai grave, pacientul nici măcar nu-și mai dă seama că există o parte stângă.

Se poate ca un pacient care suferă de neglijare hemispațială să se bărbierească doar pe o parte a feței, să se îmbrace doar pe o jumătate a corpului și să mănânce mâncarea dintr-o jumătate a farfuriei. Dacă rotești farfuria, nu va ști de unde a apărut mâncarea în plus. Dacă îi ceri să deseneze, va realiza doar partea dreaptă a desenului și nu va observa că a omis jumătate de desen. Dacă îi ceri să deseneze un ceas, de obicei va desena un cerc complet, poate pentru că păstrează o amintire motorie a mișcării circulare a mâinii. Dar apoi va înghesui numerele de la 1 la 12 în partea dreaptă a cercului și va crede că a desenat corect.[0]

Când eram la facultate, am întâlnit un pacient cu neglijare hemispațială care a venit la laboratorul nostru. De obicei nu făceam cercetare clinică, dar cunoșteam testul de bază. Am luat o foaie de hârtie și am umplut-o cu linii

[0] K.M. Heilman și E. Valenstein, „Mechanism Underlying Hemispatial Neglect”, Annual Neurology 5 (1972): 166 – 170; M. Kinsboume, „A Model for the Mechanism of Unilateral Neglect of Space”, Transactions of the American Neurological Association 95 (1970): 143 – 146; M.M. Mesulam, „A Cortical Network for Directed Attention and Unilateral Neglect”, Annual Neurology 10 (1981): 309 – 325; S.M. Szezepański, C.S. Konen și S. Kastner, „Mechanisms of Spațial Attention Control în Frontal and Parietal Cortex”, Journal of Neuroscience 30 (2010): 148 – 160.

[0] P. Chen și K.M. Goedert, „Clock Drawing în Spațial Neglect: A Comprehensive Analysis of Clock Perimeter, Placement, and Accuracy”, Journal of Neuropsychology 6 (2012): 270 – 289.

scurte, orizontale, împrăștiate aleatoriu pe toată pagina. Apoi i-am așezat-o în față și i-am cerut să taie toate liniile pe care le vedea.

Îmi amintesc că a râs și a zis: „Iar asta!” S-a uitat atent la pagină și a tăiat fiecare linie din partea dreaptă. Când l-am întrebat dacă era *sigur* că a terminat, s-a uitat atent și a spus: „Ai dreptate, am omis câteva”. Apoi a tăiat rapid alte câteva linii de la jumătatea paginii, lăsându-le neatinse pe cele din stânga. Am rotit pagina, și a fost uimit să vadă cât de multe linii nu tăiase. Pentru el, testul era o surpriză de fiecare dată. Pacienții cu sindrom de neglijare hemispațială știu că ceva e în neregulă, însă nu pot spune ce anume.

Unul dintre cele mai tulburătoare experimente legate de neglijarea hemispațială implică un test de imaginație vizuală.[0] Pacienților li s-a cerut să-și imagineze că stau în partea de nord a unei piețe dintr-un oraș familiar și să indice toate clădirile pe care și le aminteau. Le-au indicat prompt pe cele care ar fi fost în dreapta lor. Apoi li s-a cerut să-și imagineze că sunt în partea de sud a aceleiași piețe. De data asta au indicat seria opusă de clădiri. Știau că fac o greșeală, dar nu puteau să înțeleagă în ce consta. Nici nu puteau să-și imagineze o parte stângă a acelei piețe.

Și, cu toate astea, informația din partea stângă, neglijată, nu este ștearsă cu totul. Într-un experiment deosebit de relevant,[0] o pacientă cu neglijare hemispațială s-a uitat la imaginea unei case din a cărei fereastră din partea stângă izbucneau flăcări. Când i s-a cerut să descrie imaginea, pacienta a descris o casă obișnuită. Nu era conștientă de flăcări, dar a spus că nu-i plăcea casa. Ceva nu părea în regulă, chiar dacă nu putea spune ce anume. Acest experiment este doar unul din multele care arată cum informația din partea neglijată ajunge în creier, este procesată și poate afecta comportamentul în moduri subtile. Dacă înțepi pacientul cu un ac în partea neglijată a

---

[0] E. Bisiach și C. Luzzatti, „Unilateral Neglect of Reprezentațional Space”, Cortex 14 (1978): 129 – 133.

[0] J.C. Marshall și P.W. Halligan, „Blindsight and Insight în Visuo-Spațial Neglect”, Nature 336 (1988): 766 – 767.

corpului, se poate să se ferească și să știe că s-a întâmplat ceva neplăcut, dar să nu fie conștient de vreo durere specifică. Dacă îi arunci o minge din partea stângă, se poate să se ferească și să nu știe de ce. Neglijarea nu înseamnă ștergerea unei jumătăți a spațiului. Este absența conștiinței și atenției direcționate spre acea jumătate a spațiului.

În contextul teoriei schemei atenției, cea mai simplă explicație a neglijării este că mecanismul atenției, inclusiv o schemă a atenției, poate fi separat în cel puțin două câmpuri, drept și stâng. În relație cu partea dreaptă a spațiului, pacientul poate funcționa, mai mult sau mai puțin, în ceea ce privește atenția, acțiunea, cogniția socială și conștiința subiectivă. Față de partea stângă, pacientul este un zombi lipsit de o schemă a atenției.

Neglijarea hemispațială este de obicei provocată de atacuri cerebrale care duc la leziuni la nivelul unei mari regiuni a cortexului, ceea ce face dificilă identificarea regiunii corticale răspunzătoare de principalele simptome. Căutând zonele cerebrale afectate la majoritatea pacienților, specialiștii în neuroștiințe au izolat epicentrul neglijării. Vătămarea multor regiuni cerebrale poate determina diferite grade de neglijare, dar de departe cele mai severe și mai durabile cazuri sunt provocate de leziunile la nivelul JTP, în special în partea superioară, suprapusă lobului parietal.[0] Leziunea acestei regiuni critice a creierului perturbă conștiința.

Neglijarea hemispațială, oricât ar fi de devastatoare pentru pacienți, afectează totuși doar o jumătate a spațiului. Pacienții rămân în parte conștienți. Pot să interacționeze cu cei dragi și să-și trăiască viața în cealaltă jumătate a spațiului. În multe cazuri, sindromul se estompează în timp, pe măsură ce creierul se reorganizează, și pacientul își recuperează parțial, dacă nu în întregime, capacitatea pierdută.

---

[0] G. Vallar și D. Perani, „The Anatomy of Unilateral Neglect After Right-Hemisphere Stroke Lesions: A Clinical/CT-Scan Correlation Study în Man”, Neuropsychologia 24 (1986): 609 – 622.

Pacienții care suferă atacuri cerebrale au însă leziuni atât de extinse, încât își pierd conștiința cu totul. Mulți dintre acești pacienți au o responsivitate minimă. Reacționează dacă îi ciupești sau le sufli aer în ochi. Nu sunt în moarte cerebrală – mare parte a creierului rămâne intactă și funcțională. Dar nu prezintă nicio reacție mai complexă. Nu par să mai fie conștienți. Acestea sunt cazurile cele mai triste, dat fiind că esența persoanei pare să fi dispărut.

Acestor pacienți neconștienți sau aflați într-o stare vegetativă le sunt făcute adesea evaluări imagistice pentru a determina cât de grave sunt leziunile lor cerebrale. Vătămarea poate fi masivă și variază de la pacient la pacient. Dar, în mod tipic, regiunile cerebrale cel mai frecvent asociate cu pierderea conștiinței sunt rețelele parietal-frontale.[0] Eliminarea acestor rețele din ambele emisfere cerebrale înseamnă eliminarea conștiinței. Un zombi lipsit de o schemă a atenției nu este o invenție filosofică, o ipoteză sau un experiment de gândire. Este o

[0] M.A. Bruno, S. Majerus, M. Boly, A. Vanhaudenhuyse, C. Schnakers, O. Gosseries, P. Boveroux, M. Kirsch, A. Demertzi, C. Bernard, R. Hus tinx, G. Moonen și S. Laureys, „Funcțional Neuroanatomy Underlying the Clinical Subcategorization of Minimally Conscious State Patients”, Journal of Neurology 259 (2012): 1087 – 1098; S. Laureys „The Neural Correlate of (Un) Awareness: Lessons From the Vegetative State”, Trends în Cognitive Sciences 9 (2005): 556 – 559; S. Laureys, S. Antoine, M. Boly, S. Elinex, M.E. Faymonville, J. Berré, B. Sadzot, M. Ferring, X. De Tiège, P. Van Bogaert, I. Hansen, P. Damas, N. Mavroudakis.
B. Lambermont, G. Del Fiore, J. Aerts, C. Degueldre, C. Phillips, G. Franck, J.L. Vincent, M. Lamy, A. Luxen, G. Moonen, S. Goldman, și P. Maquet, „Brain Function în the Vegetative State”, Acta Neurologica Belgica 102 (2002): 177 – 185; J. Leon-Carrion, U. Leon-Dominguez.
L. Pollonini, M.H. Wu, R.E. Frye, M.R. Dominguez-Morales și G. Zouridakis, „Synchronization Between the Anterior and Posterior Cortex Determines Consciousness Level în Patients with Traumatic Brain înjury”, Brain Research 1476 (2012): 22 – 30; D. Roquet, J.R. Foucher, P. Froehlig, F. Renard, J. Pottecher, H. Besancenot, F. Schneider, M. Schenck, și S. Kremer, „Resting-State Networks Distinguish Locked în From Vegetative State Patients”, Neuroimage: Clinical 12 (2016): 16 – 22.

tragedie medicală cât se poate de reală.

7
Problema dificilă și alte perspective asupra conștiinței

E tentant să crezi că conștiința ține de opinia academică și că circulă o mulțime de păreri diverse ce formează un mare haos. Această impresie alimentează ideea comună potrivit căreia conștiința trebuie să fie vreun mister de nerezolvat. Eu nu sunt atât de pesimist. Cred că suntem aproape de înțelegerea mecanismului conștiinței și că îi cunoaștem deja principiile de bază. Un motiv pentru optimismul meu este că multe dintre teorii și opinii nu sunt atât de diferite pe cât par la o primă vedere. Dacă sapi suficient, descoperi aspecte comune și interesante.

În acest capitol voi explica modul în care teoria schemei atenției se poate lega de alte perspective științifice bine-cunoscute asupra atenției. Voi contrazice unele teorii și voi găsi puncte în comun cu altele. Nu pot oferi nicidecum o expunere completă a acestor perspective alternative într-un spațiu atât de restrâns. În loc să evaluez fiecare teorie, voi accentua modul în care se leagă de teoria schemei atenției.

PROBLEMA DIFICILĂ ȘI METAPROBLEMA

Filosoful David Chalmers a introdus termenul *problemă dificilă* (*hard problem*), influențând astfel dezbaterea științifică asupra conștiinței pentru următoarele câteva decenii.[0] Conștiința este o problemă dificilă pentru că e o experiență personală; nu poate fi confirmată din exterior. Experiența subiectivă, prin însăși natura ei, e imună la

---

[0] D. Chalmers, „Facing Up to the Problem of Consciousness”, Journal of Consciousness Studies 2 (1995): 200 – 219.

știință. Nu poți să aplici presiune și să-i măsori forța de reacție, nu poți s-o pui pe cântar ca să vezi cât cântărește și nu poți să o încălzești ca să-i măsori temperatura de ardere… „Problema dificilă” e de fapt un eufemism pentru problema *imposibilă* pe care știința nu o poate aborda niciodată.

Chalmers a venit acum cu termenul *metaproblemă*, care se referă la motivul pentru care credem că avem o problemă dificilă.[0] S-ar putea însă să nu avem nicio problemă. S-ar putea să nu avem undeva, înăuntrul nostru, o esență non-fizică, în mod fundamental inexplicabilă. Poate că sarcina noastră ca oameni de știință este să explicăm motivul pentru care oamenii tind să creadă că s-ar confrunta cu o astfel de problemă dificilă.

Teoria schemei atenției corespunde cu această a doua abordare. Dacă reduci teoria la esența ei, ea este o explicație pentru modul în care o mașinărie biologică presupune în mod greșit existența unei probleme dificile. Când mașinăria accesează schema atenției – o reprezentare simplificată, schematică, a propriilor procese interne –, este informată că deține o proprietate interioară, subiectivă, eterică a conștiinței.

Aspectul central al teoriei este puterea cunoașterii bazate pe modele în opoziție cu cunoașterea superficială. Următorul exemplu va explica mai bine ce vreau să spun.

Să presupunem că o fetiță se joacă. Latră, merge în patru labe și spune: „Sunt un cățeluș!” Pentru a face această afirmație, creierul ei trebuie să construiască propoziția-cheie: „Sunt un cățeluș”, precum și să aibă informația că aceste animale latră și merg în patru labe. Și totuși această afirmație există într-un context mai larg. Creierul conține o rețea vastă de informații, inclusiv: „În realitate nu sunt cățel”, „Inventez asta ca să mă joc”, „Sunt fetiță” și așa mai departe. Unele dintre aceste informații sunt prezente la nivel cognitiv și lingvistic. Cele mai multe există la un nivel mai profund, senzorial sau perceptiv.

[0] D. Chalmers, „The Meta-Problem of Consciousness”, The Journal of Consciousness Studies 25 (2018): 6 – 61.

Schema ei corporală este construită automat, sub nivelul superior al cogniției, și prezintă silueta fizică a unui corp uman, nu a unui cățeluș. Ea își vede mâinile de om, și informația vizuală îi confirmă identitatea umană. Își amintește că a mâncat cu lingura fulgii de porumb de la micul dejun, că a fost la școală, că a citit o carte – toate activități umane. Afirmația „Sunt un cățeluș" este o propoziție superficială, care nu corespunde celor mai profunde modele interne ale sale.

Dar să presupunem că am niște instrumente științifico-fantastice care pot manipula informația din creierul ei. Îi modific schema corporală pentru a reflecta corpul unui cățeluș. Modific informația din sistemul ei vizual și din memoria ei pentru a o face convergentă propoziției că e un cățel. Elimin informația cognitivă specifică, potrivit căreia ea spune: „Inventez asta ca să mă joc". Înlocuiesc informația: „Sunt sigură că asta nu e adevărat" cu propoziția opusă. Cum ar ști ea că nu e un cățel? Creierul ei este ostaticul informației pe care o deține. În mod tautologic, știe doar ce știe. Ea nu s-ar mai gândi că identitatea ei de cățeluș ar fi doar o ipoteză sau doar un joc. Ar considera că e adevărată. Nu ar avea niciun motiv să creadă altceva.

Ai putea încerca să o convingi că lucrurile nu stau așa. Ai putea spune: „Dar înțelegi engleza și poți să vorbești. Cățelușii nu fac asta. Nu crezi că asta înseamnă că ți-ai greșit identitatea?"

Să presupunem că e o fetiță dotată intelectual și că înțelege logica argumentului. Noua informație va fi la un nivel superficial, cognitiv. Va intra în conflict cu modelele ei interne profunde. Fetița se va afla în situația de a crede intuitiv un adevăr despre ea însăși și în același timp de a avea un adevăr intelectual diferit.

La fel, scriind această carte, s-ar putea să te conving că conștiința ta are la bază o schemă a atenției. Conform acestui argument intelectual, afirmi că ești conștient pentru că ai o serie de informații care susțin asta. Dar, în mod intuitiv, crezi altceva despre propria persoană. Dacă te

bazezi pe introspecție, dacă accesezi tocmai această schemă a atenției, ea îți va spune propria poveste. Informația pe care o deține îți spune că nu, conștiința ta nu e făcută din informații sau din vreun mecanism sau din neuroni – este o esență eterică și o caracteristică inerentă ție. Dacă am făcut o treabă bună și te am convins de argumentul meu, te vei găsi într-un conflict, cunoașterea superficială, intelectuală, îndreptându-te într-o direcție, iar modelele profunde, interne, ancorându-te într-alta. Nu vei putea niciodată să unești aceste perspective. Nu poți folosi cunoașterea superficială, intelectuală, sau câteva ore de gândire pentru a șterge o schemă a atenției care s-a format în milioane de ani de evoluție și care există în adâncul sistemului tău.

Pentru a oferi un alt exemplu de model imperfect existent undeva în adâncul nostru, să ne gândim la modul în care vedem culoarea alb. Sistemul vizual construiește un model de alb ca luminozitate în absența unor culori care să o contamineze. Acest model a evoluat în cursul a milioane de ani și este comun multor specii de animale. În cele din urmă, în anul 1671, un animal deosebit de inteligent, Isaac Newton, a descoperit că acest model este o simplificare.[0] Lumina albă este de fapt un amestec al tuturor culorilor, și creierul o reprezintă într-un mod simplificat.

Am putea spune că „problema dificilă” a culorii alb este: „Care e acel proces fizic special care purifică lumina albă de tot ceea ce o contaminează?” Metaproblema corespondentă este: „De ce credem că există o problemă dificilă? De ce credem că lumina albă este purificată?” Cunoaștem acum răspunsul la metaproblemă – creierul construiește un model simplu, pragmatic, dar imperfect – și, prin urmare, știm că nu este necesar să rezolvăm problema dificilă.

---

[0] LA. Newton, „Letter of Mr. Isaac Newton, Profesor of the Mathematics în the University of Cambridge; Containing His New Theory about Light and Colors: Sent by the Author to the Publisher from Cambridge, Febr. 6.1671/72; În Order to be Communicated to the Royal Society”, Philosophical Transactions Royal Society 6 (1671): 3075 – 3087.

Și, cu toate astea, chiar dacă orice om educat înțelege acum că lumina albă este un amestec al celorlalte culori, această cunoaștere nu schimbă modelul integrat în sistemul vizual. Vedem în continuare albul ca pe o culoare pură, și nu ca pe un amestec. Contradicția nu pare să deranjeze pe nimeni. Ne-am obișnuit cu un strat de cunoaștere intelectuală, cognitivă, aflat în conflict cu modelele profunde, înnăscute ale creierului. Se poate spune că știința este procesul treptat prin care părțile cognitive ale creierului nostru descoperă inexactitățile modelelor profunde, integrate evoluționist, pe care le avem despre lumea din jur.

Filosoful François Kammerer a pus o întrebare foarte pertinentă despre teoria schemei atenției.[0] Să presupunem că teoria este corectă. Creierul construiește o schemă a atenției care reprezintă atenția. Aceasta prezintă caracteristici generale, de nivel superior, cum ar fi capacitatea noastră de a ne concentra și de a procesa informația în profunzime. În același timp, nu cuprinde nicio descriere a proprietăților fizice sau mecanice ale atenției. Nu specifică faptul că atenția *nu are* o substanță fizică – nu are nimic de zis despre asta. Nu ne informează deloc în privința părților componente, cum ar fi neuronii și sinapsele. Dacă intuițiile noastre despre conștiință sunt formate pe baza acestui model intern, atunci de ce avem o intuiție atât de intensă că conștiința noastră este o esență eterică? De unde avem intuiția că conștiința nu are substanță sau greutate fizică dacă această informație nu este conținută în acest model intern?

Răspunsul este, cred, că în general oamenii nu au această intuiție. Nu înțelegem conștiința ca neavând substanță fizică, ci o înțelegem ca fiind ceva pentru care atributele fizice sunt *irelevante*. Iar aceste două intuiții sunt foarte diferite.

Pentru a înțelege ce vreau să spun, imaginează-ți că

[0] F. Kammerer, „The Hardest Aspect of the Illusion Problem – And How to Solve It”, Journal of Consciousness Studies 23 (2016): 124 – 139.

cineva te bate pe umăr. Această atingere activează receptorii tactili, care transmit informația către creier. În fond, creierul tău construiește un tip anume de model intern, un model tactil, un pachet de informații care descriu acea atingere. Modelul conține informații despre locul atingerii, despre intensitatea inițială, despre presiune și durată și poate chiar despre textura fină sau moale a degetului. Este o reprezentare senzorială bogată. Dar nu conține nicio informație despre gust. O atingere pe umăr nu vine la pachet cu un gust sărat, de exemplu. Nu vreau să spun că atingerea este searbădă sau fără gust și că ar avea nevoie de sare; ci că nu are *nicio* legătură cu domeniul gustului. Nu ocupă același spațiu informațional. Acum, că am pomenit această posibilitate, poți să o iei în calcul într-un sens superficial, cognitiv, dar nu poți să modifici modelul intern profund. Percepția tactilă este un proces înnăscut și nu e permeabil la modificări cognitive. Nu poți să faci așa încât o atingere să aibă gust.

Sunt destul de sigur că, dacă ai monta electrozi în creierul unui om normal și ai citi informația codificată în sistemul tactil, modelul perceptiv al atingerii *nu* ar conține informația „Și, apropo, gustul nu e prezent”. Nu e necesară negația explicită. Modelul omite pur și simplu proprietățile gustative. Nu înțelegem intuitiv atingerea ca pe ceva pentru care nu există informații despre gust, ci o înțelegem ca pe ceva pentru care gustul este *irelevant*.

Eu susțin că schema atenției funcționează în același mod. Conține un set bogat de informații, însă limitat. Descrie proprietățile generale ale atenției, dar nu proprietățile fizice, mecaniciste. Pe baza acestui model intern, credem intuitiv într-o experiență mintală dinăuntrul nostru care poate deține informații și ne poate face să acționăm, așa cum face atenția, dar care nu are o relație anume cu nimic fizic. Aspectul fizic este irelevant pentru ea. Această esență mintală, experiența în sine, este, în acest sens, *metafizică*. Nu poți să o atingi, nu e netedă, nu are textură, nu e dură, nu are relief, nu e grea, ușoară, mirositoare, verde sau ascuțită. Nu se încadrează nicăieri în

aceste dimensiuni fizice, după cum atingerea nu există în dimensiunea gustului.

Și totuși, conform acestei teorii, schema atenției face referire la cel puțin o proprietate fizică. Prezintă atenția ca având un centru fizic undeva în noi. Cu acest tip de model intern ar trebui să avem o intuiție despre o esență mintală suprapusă lumii fizice, în sensul că putem să indicăm un loc și să spunem: „E cam pe-aici”. E un fel de duh care se află într-un spațiu fizic, deși îi lipsesc alți parametri fizici. Potrivit acestei teorii, fantoma din mașină (*ghost în the machine*), energia conștientă din noi, este o intuiție provenită direct din schema atenției, cu o descriere incompletă a atenției.

Și așa ajungem înapoi la problema dificilă și la metaproblemă. Problema dificilă derivă din presupunerile provenite din acel model profund, situat în adâncime, numit schema atenției. Teoria schemei atenției este meta răspunsul care explică de ce oamenii cred într-o problemă dificilă.

ILUZII ȘI METAFORE

Este conștiința o iluzie?

Iluzionismul este o abordare a conștiinței relativ nouă, aflată încă în dezvoltare. Ideea centrală este că de fapt nu avem conștiință. Experiența în sine, esența subiectivă, e absentă. În schimb, o iluzie creată de creier ne face să *credem* că suntem conștienți. Poate această iluzie are un avantaj funcțional specific (să ofere vieții mai mult gust estetic, conform uneia dintre ipoteze) sau poate nu are niciun rol în supraviețuire și este doar o consecință întâmplătoare a modului în care creierul procesează informația.[0] Vom înțelege cândva mecanismul din spatele ei și atunci vom ști dacă are vreo semnificație funcțională.

Până atunci, oamenii de știință nu mai au de ce să

[0] Pentru mai multe informații despre argumentul conform căruia conștiința este o iluzie care sporește satisfacția vieții: N. Humphrey, Soul Dust (Princeton, NJ: Princeton University Press, 2011).

explice cum ia naștere o conștiință neexistentă, așa cum nu trebuie să explicăm de ce Pământul este plat sau de ce Soarele orbitează în jurul Pământului. Susținătorii acestei perspective câștigă din ce în ce mai mult teren în domeniul științific, dar ideea e adesea greu de acceptat pentru publicul larg.

Teoria schemei atenției este un fel de iluzionism. Conform acestei teorii, cea mai de neînțeles însușire a conștiinței – natura sa eterică, metafizică – nu este reală. Credem că o astfel de esență există doar pentru că suntem induși în eroare de un model intern imperfect.

Însă, în experiența mea, să numești conștiința o iluzie înseamnă să-i dai teoriei sărutul morții. S-ar putea ca o mână de filosofi să înțeleagă ce vrei să spui, dar restul lumii o să-ți respingă teoria ca pe o prostie elitistă: „Cum să fie conștiința o iluzie, când e evident că în capul meu se petrec o grămadă de lucruri?”

Cuvântul *iluzie* este atât de alunecos, încât poate îngreuna discuțiile despre conștiință. Voi prezenta aici trei greșeli care se fac îndeobște atunci când se vorbește despre conștiință ca despre o iluzie – fără să atac fundamentele abordării iluzioniste, pe care o consider în esență corectă.

Am văzut un băiețel și o fetiță care se jucau în nisip pe plajă. Cred că aveau în jur de cinci ani. Băiețelul a spus foarte candid:

— Nu trebuie să stăm prea mult la soare, ca să nu ne crească clești.

Fetița era uimită.

— Serios? a întrebat holbându-se la el.

Băiețelul a dat solemn din cap:

— Așa e.

Și-a ridicat mâinile și a făcut o mișcare ca și cum ar fi strâns din clești.

— Mi-a zis mama.

Episodul este fermecător pentru că e evident că băiatul a înțeles greșit o metaforă obișnuită. Mama lui trebuie să-i

fi spus că o să se facă roșu ca racul – adică o să se ardă de la soare. Mintea lui s-a îndreptat însă spre clești, nu spre culoarea roșie.

Metaforele respectă reguli stricte, implicite.[0] Într-o metaforă obișnuită, o entitate (racul) este comparată cu o altă entitate (omul ars de soare). Doar un atribut este relevant. Racul are multe însușiri – clești, exoschelet, ochi compuși situați pe antene –, dar e de așteptat ca o persoană care aude metafora să înțeleagă care atribut anume, esențial, trebuie transferat. Cu toții folosim metaforele astfel.

Când un om de știință spune: „Conștiința este o iluzie”, cred că majoritatea oamenilor consideră această afirmație în mod implicit o metaforă. Baza metaforei, iluzia vizuală, are multe aspecte posibile. Un obiect poate părea mai mare, sau mai înclinat, sau mai departe decât e de fapt. Un obiect fix poate părea în mișcare. O suprafață convexă poate părea concavă. Dar când cuvântul *iluzie este* folosit în contextul unei metafore înseamnă de obicei un singur lucru. Oamenii izolează o calitate esențială: echivalează „iluzia” cu „mirajul”. În cazul mirajului, crezi că e prezent un lucru care de fapt nu e acolo. Nu greșești doar dimensiunea sau detaliile – înțelegi greșit însăși existența acelui lucru.

De exemplu, să presupunem că un prieten ți se plânge: „Jur, competența șefului meu e o iluzie”. Acea persoană nu vrea să spună: „Este un șef competent, dar într-un mod puțin diferit decât te-ai aștepta. De fapt, s-ar putea să fie *mai* competent decât te-ai aștepta”. Nu, prietenul tău vrea să spună că șeful lui nu e deloc competent. În contextul unei metafore, să numești ceva o iluzie înseamnă să negi fiecare aspect al existenței sale.

Atunci când afirmă că conștiința este o iluzie, majoritatea oamenilor nu înțeleg prin asta: „Conștiința se aseamănă unei iluzii vizuale pentru că procesarea informației care se desfășoară în mintea noastră este puțin

[0] S. Glücksberg, Understanding Figurative Language (Oxford, UK: Oxford University Press, 2001).

diferită de ceea ce susținem pe baza introspecției pornind de la un model intern imperfect – și toate astea seamănă cu o iluzie vizuală”. Oamenii înțeleg cu totul altceva, și anume că în spatele iluziei nu există nimic. Nu există nimic acolo. Conștiința nu există, și punct – afirmație complet absurdă pentru majoritatea oamenilor.

După cum am menționat în capitolele anterioare, conform teoriei schemei atenției, conștiința nu este nicidecum un miraj. E o expunere simplificată, imperfectă, a ceva real. Creierul chiar asimilează informații și le procesează în profunzime. Când afirmăm că avem o experiență subiectivă, oferim o versiune ușor schematizată a adevărului literal. Adică există ceva acolo. Conștiința seamănă cu o iluzie în sens practic, și acesta este motivul pentru care teoria schemei atenției este efectiv o teorie iluzionistă. Dar conștiința nu este o iluzie în sensul în care majoritatea oamenilor înțeleg această metaforă.

Să aprofundăm discuția despre iluzii și despre felul în care diferă ele de percepția normală. Sistemul tău vizual construiește modele interne, reprezentări simplificate ale obiectelor de peste tot din jurul tău. Face asta automat și constant atâta timp cât ai ochii deschiși. Simplul fapt de a vedea nu este, în sine, o iluzie. Iluzia este un caz mai degrabă specific; apare când sistemul face o greșeală. Un obiect mic pare mare sau ceva drept pare înclinat. Undeva apare o eroare. Majoritatea specialiștilor în sistemul vizual înțeleg cuvântul *iluzie în* sensul unei erori sau abateri de la normal.

Schema atenției pe care o propun este un model intern normal, construit automat și continuu. Îți oferă informații simplificate despre starea atenției tale. Presupun că ar putea să funcționeze greșit, caz în care eroarea din modelul intern ar constitui o iluzie. Să crezi că o păpușă are conștiință este o iluzie. Fiecare model intern este o versiune simplificată a realității, realitatea fiind mult mai complexă și mai fin texturată decât capacitatea și nevoia de procesare a creierului. Dacă orice model intern

simplificat se califică drept iluzie, atunci termenii *vizual* și *iluzie* sunt redundanți, dat fiind că orice imagine vizuală este o iluzie. Dacă aceasta e noua definiție a cuvântului, atunci tot ce vedem, auzim, simțim și gândim este iluzie. Poate că filosofii se gândesc la această definiție extinsă atunci când numesc conștiința iluzie – conștiința este parte a modului imperfect în care creierul înțelege realitatea. Înțeleg logica acestui argument, dar rămân sceptic în privința acestei definiții atât de extinse. Cuvântul își pierde sensul dacă se aplică la orice.

Poate că cel mai mare pericol care decurge din a numi conștiința iluzie este că oamenii riscă să înțeleagă greșit această idee ca pe un raționament circular. Pentru majoritatea oamenilor, o iluzie este prin definiție un tip de experiență conștientă. Dacă conștiința este o iluzie, atunci cine construiește iluzia? Raționamentul pare să folosească conștiința pentru a explica conștiința.

Această critică poate fi frustrantă pentru că pornește dintr-o neînțelegere. Pentru iluzioniști, nimic din creier nu resimte iluzia conștiinței. În schimb, creierul susține că are conștiință pe baza unor informații imperfecte.

Mi se pare interesant să vorbesc cu filosofi care spun, ca și cum ar fi uimiți de întrebare: „*Bineînțeles* că nimic din creier nu are o experiență subiectivă a iluziei că posedă conștiință. Nu asta am vrut să spun. Conceptualizarea ar fi circulară dacă ar fi așa!” După care vorbesc cu altcineva din orice alt domeniu care spune cu aceeași uimire: „Dar *bineînțeles* că iluzia implică existența unei conștiințe care să o resimtă! Ce altceva înseamnă acest cuvânt? De ce să folosești cuvântul ăsta dacă te referi la altceva? De ce nu numești conștiința «liliac frugivor» și nu-ți inventezi propria limbă?”

Așa că nu numesc teoria schemei atenției o teorie iluzionistă, chiar dacă, potrivit iluzioniștilor, chiar asta este. Dezacordul dintre noi este mai degrabă de natură terminologică. Poate conștiința funcționează ca o caricatură, dat fiind că distorsionează ceva real. Nu sunt însă deloc sigur că e mai clar dacă spun că conștiința este o

caricatură, în loc să spun că e o iluzie. Probabil că o formulare memorabilă sau un slogan despre conștiință nu o să fie niciodată pe deplin satisfăcătoare din punct de vedere intelectual.

O mașinărie incredibil de complicată și de subtilă are o cunoaștere bazată pe modelul propriei existențe și al lumii. Un anumit model intern, schema atenției, ne oferă intuiții despre conștiință. Cele mai multe însușiri pe care le asociem cu conștiința există concret în creier sub forma nivelurilor superioare ale atenției corticale. Unele caracteristici ale conștiinței, cum ar fi natura sa metafizică, eterică, sunt rezultatul informațiilor incomplete sau schematice din acel model intern.

Este conștiința o iluzie datorită acestor proprietăți? Poți să o numești așa dacă definești atent cuvintele. Cred că prietenii mei filosofi de orientare iluzionistă au o abordare inteligentă și folosesc definiții precise. Am intenția sinceră de a nu-i contrazice. De fapt, speranța mea este doar să demonstrez cât de bine pot funcționa perspectivele noastre în același cadru de referință.

MEMBRE-FANTOMĂ

Dacă o să ai vreodată ghinionul să-ți pierzi o mână sau un picior, o să ai probabil experiența membrului-fantomă, pe care-l simți ca și cum ar fi în continuare acolo. Aproximativ 90% dintre oamenii care au suferit o amputare au o senzație-fantomă cel puțin temporar, iar pentru unii pacienți aceasta poate dura ani întregi.[0] Senzațiile de mișcare a încheieturii, de atingere, de durere, de frig și căldură, toate se păstrează. Vezi că membrul nu mai există, dar simți în continuare fiecare parte din el. Asemenea unei fantome, membrul se extinde invizibil din corp. Lordul

[0] V.S. Ramehandran și W. Hirstein, „The Perception of Phantom Limbs”, Bra în 121 (1998): 1603 – 1630; A. Woodhouse, „Phantom Limb Sensation”, Clinical and Experimental Pharmacology and Physiology 32 (2005): 132 – 134.

Nelson, marele amiral britanic care și-a pierdut o mână în lupta de la Santa Cruz de Tenerife, a făcut afirmația celebră că putea să dovedească existența vieții de dincolo pentru că, dacă mâna lui devenise o fantomă, la fel urma să se-ntâmple și cu restul corpului său.[0]

Membrele-fantomă nu sunt doar curiozități medicale. Experiența este extrem de neplăcută. Imaginează-ți că ai un membru care pare real din punct de vedere psihologic, dar nu poți să scarpini o mâncărime, să îndoi o încheietură sau să scapi de durere. Membrul-fantomă poate fi extrem de dureros și debilitant, precum și bizar și derutant.[0] De exemplu, uneori membrul-fantomă poate să „se micșoreze” și pacientul afirmă că o mână-fantomă îi iese din umăr sau că un picior-fantomă s-a scurtat și nu mai ajunge la podea.

Explicația acceptată în general pentru membrele-fantomă este că, deși membrul în sine nu mai e, modelul intern din creier este în continuare prezent.[0] O serie bogată de informații care descriu membrul rămân în circuitele cerebrale. Acest fenomen arată cât de puternic poate fi un model intern. Dacă văzul nu ar dovedi constant nonexistența membrului, nu ai avea cum să știi adevărul. Ai presupune că încă ai acel membru.

Lucrurile se întâmplă tocmai invers în cazul sindromului clinic numit somatoparafrenie.[0]12 Pacienții care suferă leziuni provocate de atacuri cerebrale la nivelul lobului parietal pierd modelul intern al unui membru. Au acel membru, pot să îl vadă, dar nu pare să le aparțină. Neurologul Oliver Sacks, care avea un talent special de a descrie cazuri clinice cu multă empatie, a povestit despre un bărbat care își pierduse reprezentarea unui picior. Sacks

---

[0] V.S. Ramehandran și W. Hirstein, „The Perception of Phantom Limbs”, Brain 121 (1998): 1603 – 1630.

[0] Y. Luo și T.A. Anderson, „Phantom Limb Pain: A Review”, International Anesthesiology Clinics 54 (2016): 121 – 139.

[0] V.S. Ramehandran și W. Hirstein, „The Perception of Phantom Limbs”, Brain 121 (1998): 1603 – 1630.

[0] G. Vallar și R. Ronchi, „Somatoparaphrenia: A Body Delusion. A Review of the Neuropsychological Literature”, Experimental Brain Research 192 (2009): 533 – 551

povestește următorul dialog:

> *„Potoliți-vă!” am zis. „Fiți calm! Nu vă enervați! Eu n-aș lovi așa piciorul ăla”.*
> *„De ce nu!” a întrebat iritat, arțăgos.*
> *„Pentru că este piciorul dumneavoastră”, am răspuns. „Nu vă recunoașteți propriul picior?”* [0]

Una dintre cele mai grăitoare demonstrații ale modelului intern al corpului provine din studiile care au folosit mâini de cauciuc. Prima demonstrație sistematică a iluziei mâinii de cauciuc a fost publicată în 1998 de Matthew Botvinick și Jonathan Cohen.[0] Am participat la o versiune ulterioară a experimentului în laboratorul lui Henrik Ehrsson din Stockholm. Am stat pe un scaun în fața unei mese și mi-am băgat mâna într-o gaură din lateralul unei cutii de pantofi. Pe cutie era o mână de cauciuc de culoarea pielii, puțin prea mică pentru mine și nu tocmai realistă. M-am abținut să nu chicotesc.

Cercetătorul mi-a pus un inel pe arătătorul care era în cutie. Inelul avea o tijă scurtă de plastic care trecea printr-o gaură mică din partea de sus a cutiei și se atașa la un alt inel de pe arătătorul mâinii de cauciuc. De fiecare dată când îmi mișcăm arătătorul în cutie, prin cuplare mecanică se mișca și arătătorul mâinii de cauciuc.

Nu a fost nevoie decât de cinci sau șase mișcări de deget pentru ca iluzia să mă lovească din plin. Dintr-odată, mâna de cauciuc a devenit mâna mea. Senzația a fost șocantă și profundă. Schema mea corporală încorporase o bucată de cauciuc despre care știam cognitiv că nu avea nimic de-a face cu mine. E greu de descris în cuvinte cât de ciudat m-am simțit. Schema corporală nu echivalează cu informația vizuală despre o mână. Nu e cunoaștere intelectuală. Nu e cunoaștere medicală despre părțile conectate la corp. Nu e o poveste pe care mi-o spun singur.

---

[0] Oliver Sacks, Omul care își confunda soția cu o pălărie, Humanitas, București, 2005, p. 76.

[0] M. Botvinick și J.D. Cohen, „Rubber Hand «Feels» What Eye Sees”, Naturell (1998): 756.

E un model intern construit automat, mult sub nivelul cogniției, dincolo de domeniul voinței. Furnizează informații creierului cognitiv, și suntem captivi acestor informații. Dacă îmi spune că o mână de cauciuc este mâna mea, atunci bum! — am senzația viscerală că așa e, chiar dacă, paradoxal, știu că nu e adevărat.

Schema corpului conține informații de care creierul tău are nevoie pentru a controla mișcările. Îți spune ce obiecte aparțin corpului tău. Conține forma fiecărei părți a corpului, structura articulată a membrelor, dimensiunea și modul în care mâna se mișcă și se rotește. Ce nu conține este informație despre structura mecanicistă, detaliată, dinăuntrul corpului. Schema mâinii nu spune nimic despre structura osoasă sau îmbinarea tendoanelor, despre fibrele musculare cu contractare rapidă sau lentă, despre căile circulației sangvine sau despre moleculele proteice din celulele musculare care determină contractarea. Dacă în acest moment închizi ochii și vorbești despre mâna ta – nu despre cunoașterea ta medicală a modului în care ar trebui să fie o mână, ci despre felul în care știi acum că e mâna ta –, nu poți descrie decât proprietățile superficiale, nu detaliile mecaniciste. Descrierea superficială este oferită de schema ta corporală.

Dintr-o perspectivă filosofică, ce e un membru-fantomă? Adică ce este fantoma în sine? Cum ar trebui s-o clasificăm? Nu e un obiect. Nu e un câmp de energie. Nu iese nimic din ciot. *Pare* o esență invizibilă, forța vitală rămasă după ce au dispărut osul și carnea. E ca o fantomă, așa cum a spus Lordul Nelson.

Credința răspândită în fantome ar putea proveni din modele interne profunde, cum ar fi schema corporală, care alimentează acest tip de construct în cogniția noastră superioară. Și cu toate astea nu putem să dăm deoparte ideea de membru-fantomă ca pe o superstiție neștiințifică. Se întâmplă ceva important. Un proces util, care are loc încă dinainte ca membrul să fie amputat, e demascat doar după ce membrul nu mai e. Fantoma este o simulare pe

care o avem cu toții și, în cazul amputării, simularea se menține după pierderea membrului pe care ar trebui să îl reprezinte.

Văd aici o analogie între un membru-fantomă și conștiință – între schema corporală și schema atenției. Una este fantoma din corp și cealaltă este fantoma din cap. Ambele sunt simulări. Aparțin capetelor opuse ale aceluiași obiect – un model multipartit al sinelui. Schema corporală este un model al sinelui fizic și al modului în care acesta funcționează, pe când schema atenției este un model al unei alte părți a sinelui, al neuronilor interconectați din craniu și al modului în care *aceștia* funcționează. Ambele modele omit informații mecaniciste inutile. Sunt aproximative, superficiale și, cu toate astea, indispensabile. Se poate spune că schema atenției este o extensie specializată a schemei corporale.

Am petrecut ani întregi studiind schema corporală, nu doar modul în care creierul creează un model al corpului în sine, ci și felul în care creează un model al unui spațiu de siguranță din jurul corpului, care se ondulează și se adaptează în funcție de mișcările membrelor și capului, ca un strat gros de jeleu invizibil.[0] Faptul că am înțeles că creierul creează modele ale sinelui în modalități simplificate, nerealiste, dar profund utile m-a condus spre o teorie a conștiinței bazată pe schema atenției.

Specialistul în neuroștiințe Olaf Blanke și filosoful Thomas Metzinger oferă dovezi foarte puternice care susțin o conexiune între schema corporală și conștiință.[0] Potrivit acestora, autocunoașterea corporală este o formă primordială, minimă, de conștiință, o cunoaștere a sinelui ca agent separat de restul lumii. Evident, această înțelegere se potrivește de minune cu teoria schemei atenției.

---

[0] M.S.A. Graziano, The Spaces Between Us: A Story of Neuroscience, Evolution, and Human Nature (Oxford, UK: Oxford University Press, 2018).

[0] O. Blanke și T. Metzinger, „Full-Body Illusions and Minimal Phenomenal Selfhood”, Trends în Cognitive Sciences 13 (2009): 7 – 3.

SPAȚIUL GLOBAL DE LUCRU ȘI KRAKENUL CONȘTIINȚEI

În liceu aveam o gașcă de puști populari. Ei erau și cei care răspândeau bârfele. Circulau în școală o grămadă de zvonuri și frânturi de informații, dar dacă ceva ajungea la gașca populară devenea imediat disponibil și toată lumea afla noutatea. Așa funcționează și teoria spațiului global de lucru a conștiinței. Anumite informații din creier ajung într-un spațiu global de lucru. De aici sunt transmise peste tot și pot să ghideze comportamentul și discursul nostru.

Teoria spațiului global de lucru a fost propusă pentru prima dată de Bernard Baars în anii 1980.[0] De atunci a fost elaborată de mulți alții, în special de Stan Dehaene, și a fost racordată la ce se știe în prezent despre rețelele din cortexul cerebral.[0] Conform acestei teorii, despre care am vorbit pe scurt în Capitolul 4, informațiile trec printr-o ierarhie corticală. Unele pot atinge cea mai mare putere a semnalului, depășind celelalte semnale și ajungând la cele mai înalte niveluri de procesare, probabil în rețelele parietal-frontale. Aici, informația a ajuns la gașca bârfitoare din creier, spațiul global de lucru. A ajuns să fie celebră în creier, după cum spune Dennett.[0] Informația care ajunge în spațiul global de lucru ajunge și în conștiință.

Provocarea teoriei spațiului global de lucru constă în faptul că nu explică de ce informația, după ce ajunge în spațiul global de lucru, capătă calitatea de experiență conștientă. E o teorie incompletă – deși, într-un sens, faptul că e incompletă funcționează conform planului. Poți să analizezi substratul anatomic al conștiinței fără să trebuiască să iei poziție într-o dezbatere filosofică. Sub

[0] B.J. Baars, A Cognitive Theory of Consciousness (Cambridge, UK: Cambridge University Press, 1988).
[0] S. Dehaene, Consciousness and the Brain (New York: Viking Press, 2014).
[0] D. Dennett, Sweet Dreams (Cambridge, MA: MIT Press, 2005).

unele aspecte, seamănă cu teoria științifică inițială a conștiinței, formulată de Hipocrate acum 2.500 de ani, potrivit căreia mintea ține de creier.[0] Teoria lui Hipocrate a fost o piatră unghiulară în știință pentru că a izolat corect substratul, chiar dacă nu a explicat ce e conștiința sau cum ia naștere.

Teoria schemei atenției oferă o modalitate de a întregi tabloul. Să presupunem că teoria spațiului global de lucru e corectă și explică bine atât cât explică. Informațiile pot fi amplificate și selectate de atenție până când ajung să aibă un efect global la nivel cerebral. Teoria schemei atenției afirmă că, în plus, creierul construiește un model schematizat al atenției. Creează propria teorie metafizică naivă a ceea ce este un spațiu global de lucru.

Ce relevanță are acest lucru pentru relația dintre spațiul global de lucru, schema atenției și conștiință? Am încercat să mă gândesc la o analogie potrivită pentru a explica nuanțele și cea mai bună pe care am găsit-o este comparația dintre kraken și calamarul uriaș. Krakenul este o creatură fantastică din mitologia nordică, prima menționată în saga islandeză din secolul al XIII-lea, *Örvar-Odds*. Krakenul este un calamar gigantic cu o putere și ferocitate supranaturale, căruia îi cam place să distrugă corăbii. Deși nu există krakeni, calamari gigantici există. Trăiesc în adâncimile oceanului și pot ajunge să măsoare 15 metri. Sunt apariții rare, nu se știu multe despre ei și nu atacă niciodată corăbii, dat fiind că ar muri din cauza diferenței de presiune a apei dacă s-ar apropia de suprafață. Mitul krakenului este aproape sigur o interpretare a calamarului gigantic.

Atenția poate fi asemănată cu calamarul gigantic – un fenomen real, chiar dacă eluziv. Prezentarea nivelului cel mai înalt al atenției ca spațiu global de lucru poate fi comparată cu descrierea capului calamarului gigantic drept o pungă care conține organele – o modalitate utilă pentru oamenii de știință de a gândi anatomia acestuia. Dar

[0] C.G. Gross, Brain, Vision, Memory: Tales în the History of Neuroscience (New York: Bradford Books, 1999).

niciuna dintre acestea nu explică conștiința – krakenul. Pentru a explica krakenul și semnificația lui nu e suficient să diseci un calamar și să spui „Iată, ăsta e krakenul”. Pentru că nu e asta. Krakenul este o versiune supranaturală, distorsionată, a calamarului. Are un efect cultural și afectiv pe care un biet calamar nu l-ar putea avea niciodată. O teorie completă a krakenului ar trebui să cuprindă întrucâtva înțelegerea calamarului, dar trebuie să includă și înțelegerea procesului de mitologizare care duce la credința în kraken. Exact la fel, o explicație completă a conștiinței nu se poate opri la atenție și la spațiul global de lucru. Trebuie să încorporeze și acel model naiv al sinelui, schema atenției, care ne șoptește despre krakenul conștiinței.

GÂNDIREA DE NIVEL SUPERIOR

Exemplarul din cartea mea pe care îl citești acum are probabil un titlu pe copertă, o pagină de titlu și o descriere pe spate. Strict vorbind, toate lucrurile astea nu sunt conținut. Sunt informații *despre* conținut. Sunt metainformații care indică și etichetează. Pentru ca această carte să ajungă în lume și să provoace un efect, trebuie să-i fie atașate aceste metainformații de nivel superior.

Filosoful David Rosenthal a sugerat că ceva foarte asemănător se întâmplă cu informațiile din creier.[0] Când mă uit la un măr, ca să spun „Sunt conștient de măr”, nu e suficient ca sistemul meu vizual să proceseze informațiile despre obiect. Un gând de nivel superior trebuie să fi fost generat undeva pe parcurs și să se fi atașat informațiilor despre măr. Această propunere se numește teoria gândirii

---

[0] R.L. Gennaro, Consciousness and Self Consciousness: A Defense of the Higher Order Thought Theory of Consciousness (Philadelphia, PA: John Benjamin’s Publishing, 1996); H. L-au și D. Rosenthal, „Empirical Support for Higher-Order Theories of Consciousness”, Trends în Cognitive Sciences 15 (2011): 365 – 373; D. Rosenthal, Consciousness And Mind (Oxford, UK: Oxford University Press, 2006).

de nivel superior și aparține unui cadru general numit uneori „gândire despre gândire” sau metacogniție.[0]

Oamenii de știință continuă să dezbată ce alte informații de nivel superior mai sunt adăugate informațiilor despre măr pentru ca noi să fim conștienți de el. Poate că e vorba de informații de îndosariere.[0] Un fișier de computer conține informația primară – principalele conținuturi ale fișierului –, precum și informații de nivel superior, de îndosariere, afișate ca icoană pe desktop. Icoana reprezintă fișierul. Este o versiune simplă, comprimată, a fișierului, din care sunt excluse detaliile. Poate că sistemul nostru vizual procesează mărul și apoi comprimă informația într-un fel de icoană care devine disponibilă pentru accesarea cognitivă și raportarea verbală. Din această perspectivă, spunem că suntem conștienți de măr, pentru că, pentru cogniția noastră superioară, icoana este cumva interpretată ca act de conștiință.

O altă explicație a informațiilor de nivel superior este o evaluare a nivelului de încredere.[0] Să spunem că treci repede pe lângă un bol de fructe în care e un măr. Dacă nu ești sigur de ce ai văzut, nu vei susține că ești conștient de măr. Însă, dacă ai un nivel ridicat de încredere, vei spune că ești conștient de măr. Poate conștiința vizuală este informație vizuală plus un nivel înalt de încredere că posezi într-adevăr acea informație.

O a treia posibilitate, subliniată de filosoful Daniel Dennett, este că gândirea de nivel superior adaugă un strat

---

[0] P. Carruthers, „How We Know Our Own Minds: The Relationship Between Mândreading and Metacognition”, Behavioral and Brain Sciences 32 (2009): 12 – 182; A. Pasquali, B. Timmermans și A. Cleeremans, „Know Thyself: Metacognitive Networks and Measures of Consciousness”, Cognition 117 (2010): 182 – 190; D.M. Rosenthal, „Consciousness, Content, and Metacognitive Judgments”, Consciousness and Cognition 9 (2000): 203 – 214.

[0] D.D. Hoffman, „The Interface Theory of Perception”, în S. Dickinson.

[0] P. Grimaldi, H. L-au și M.A. Basso, „There Are Things That We Know That We Know, and There Are Things That We De Not Know We De Not Know: Confidence în Decision-Making”, Neuroscience and Biobehavioral Reviews 55 (2015): 88 – 97.

de idei mai complicat, dobândit prin cultură.[0] Cu toții suntem produse ale culturii, și poate că unul dintre cele mai răspândite mituri culturale este că avem un suflet care are experiență subiectivă. Dacă am fi fost crescuți într-o cultură radical diferită, s-ar putea să nu fi avut un construct numit conștiință, și atunci eu n-aș fi scris această carte. Dar, datorită unei idei care s-a născut poate în urmă cu zeci de mii de ani într-o peșteră sau în jurul unui foc, și care s-a răspândit în rândul întregii populații umane, am asimilat cu toții conceptul de conștiință, și gândirea noastră de nivel superior atașează acest concept oricărui lucru pe care îl facem. Ideea potrivit căreia conștiința ar fi un complex de *meme* culturale a fost propusă de psihologul Susan Blackmore.[0]

Teoria schemei atenției este o teorie a gândirii de nivel superior. Este evident că intră în aceeași categorie. Însă diferă întrucâtva de alte teorii ale gândirii de nivel superior. Conform teoriei schemei atenției, atunci când creierul construiește conștiința existenței unui măr, informația vizuală despre măr se leagă de alte informații care i se adaugă. Dar aceste informații adăugate nu aparțin întru totul „nivelului superior”. Într-un sens, sunt la același nivel cu informația despre măr. Ambele sunt reprezentări ale unor obiecte reale. Așa cum creierul construiește o reprezentare a formei mărului, a culorii mărului și a relației spațiale dintre tine și măr, construiește și o reprezentare a concentrării atenției tale asupra mărului. Această schemă a atenției nu e de natură intelectuală, conceptuală sau cognitivă. În acest sens, nu e de nivel secundar sau superior. Nu e dobândită cultural și niciun fel de educație nu ne-ar putea face să o învățăm sau să o uităm. Nu putem alege să o accesăm sau nu. A evoluat pe parcursul a milioane de ani, cu mult înaintea speciei noastre, anterior apariției vorbirii sau a cogniției umane. E construită sub

[0] D.C. Dennett, Consciousness Explained (Boston, MA: Little, Brown, and Co, 1991).

[0] S.J. Blackmore, „Consciousness în Meme Machines”, Journal of Consciousness Studies 10 (2003): 19 – 30.

nivelul limbajului, deși avem oarece acces cognitiv la ea și putem vorbi despre ea. La fel ca forma, culoarea, mișcarea sau așezarea spațială, relația atențională dintre mine și măr este o componentă a unui dosar mai mare pe care creierul îl compilează pentru măr și pentru contextul său în lume. Componenta conștiinței este, sub unele aspecte, la fel de bazală ca orice altă componentă. Credem că suntem conștienți în cel mai profund, intuitiv și nerațional mod, așa cum credem orice – pentru că creierul construiește automat modele ale lumii și ale propriei existențe și are acces cognitiv parțial la aceste modele.

Sunt de acord, desigur, și cu ideea potrivit căreia creierul creează *și* gânduri diferite, de nivel superior. Facem asta în privința culorii, de exemplu. Avem asociații afective, culturale și chiar politice cu culoarea roșie. Aceste conotații adăugate vin peste un model bazal, automat, al culorii roșii, construit în profunzimile sistemului vizual. În același mod, avem mituri culturale și personale despre conștiință – de unde vine, ce funcție îndeplinește, cum se leagă de filosofie și spiritualitate și ce se întâmplă cu conștiința după ce murim. Dar, conform teoriei schemei atenției, sub stratul asociațiilor culturale și conceptuale, care pot varia de la persoană la persoană, avem și o schemă înnăscută a atenției care e mai mult sau mai puțin la fel la toți oamenii, și care ne oferă un numitor comun al conștiinței.

ATENȚIA ȘI CONȘTIINȚA

În 1890, William James, unul dintre pionierii psihologiei moderne, scria: „Toată lumea știe ce este atenția. E asimilarea mintală într-o formă clară și intensă a unui obiect sau gând din mai multe disponibile simultan. Esența ei constă în concentrarea conștiinței. Implică renunțarea la unele lucruri pentru a se ocupa eficient de altele”.[0]

Citatul din James, celebru și bine formulat, prezintă o

[0] W. Fames, Principles of Psychology (New York: Henry Holt and Company, 1890).

confuzie tipică între atenție și conștiință. Conform expunerii sale, oamenii au ceva numit „minte” sau „conștiință”, și atenția se referă la concentrarea acesteia. Prin urmare, atenția constă în conștiință. Este focalizarea câmpului mai larg al conștiinței.

Cred că perspectiva lui James este modul în care oamenii înțeleg în mod obișnuit cuvântul *atenție*. Dar nu prea seamănă cu sensul modern, științific, al cuvântului, și nici cu modul în care-l folosesc eu în această carte. Distanța dintre sensul general și semnificația științifică a atenției este probabil principala sursă de confuzie și dezacord din jurul teoriei schemei atenției. Dată fiind această ambiguitate, nu mi-a plăcut niciodată să folosesc cuvântul *atenție*, dar are rădăcini atât de adânci în psihologie și neuroștiințe, încât până acum nu am găsit o alternativă mai bună.

În neuroștiințe, atenția este un proces cerebral care constă în faptul că o reprezentare (cum ar fi reprezentarea vizuală a unui măr) este amplificată, în timp ce semnalul altor reprezentări este redus, semnalele mai puternice având un efect mai mare asupra sistemelor cerebrale. Atenția nu este doar concentrare asupra unui obiect central; poate fi extinsă și distribuită. Dacă te gândești că ești conștient de ceva din afara atenției tale – te ocupi de A și în același timp ești marginal conștient de B, C și D –, această intuiție probabil nu e corectă; sau te bazezi pe definiția profană a atenției. Conform definiției științifice, ești, probabil, atent la toate aceste obiecte în oarecare măsură și treci rapid de la unul la altul. Potrivit sensului comun, putem considera atenția o subcategorie a conștiinței, dar, din punct de vedere științific, relația dintre cele două este cu totul alta. Atenția este o serie de mecanisme suprapuse – o metodă de procesare a datelor –, în timp ce conștiința este o experiență interioară pe care pretindem că o avem. Atenția e ceva ce face creierul; conștiința e ceva ce creierul spune că are.

În 1890, James nu putea ști încă nimic despre tehnica procesării informației. Alan Turing a descoperit principiile

mașinilor de calcul abia în anii 1930,[0] iar Claude Shannon avea să creeze teoria informației în anii 1940.[0] James nu ar fi avut cum să se gândească la atenție ca acțiune a unei mașini de calcul. El înțelegea atenția ca stare a conștiinței, și cele două au rămas asociate în mintea oamenilor de știință pentru următorii o sută de ani.

Distincția dintre atenție și conștiință a fost stabilită convingător pentru prima dată în 1999 de un grup de cercetători din Londra, printre care Robert Kentridge, Charles Heywood și Larry Weiskrantz, care au studiat un bărbat a cărui conștiință vizuală era grav afectată.[0]

Când avea opt ani, pacientul GY și-a pierdut aproape în întregime cortexul vizual primar în urma unui accident de mașină. După accident a rămas orb în partea dreaptă a câmpului vizual și în cea mai mare parte a câmpului vizual stâng. La vârsta adultă, când GY a fost așezat în fața unui ecran pe care un punct se aprindea în diferite locuri, el nu putea să-l vadă decât într-o porțiune îngustă a ecranului, puțin la stânga față de centru. Nu avea conștiință vizuală în afara acelei regiuni. Dar, când i se cerea să indice punctul luminos în câmpul vizual presupus a fi orb, în mod surprinzător putea să facă asta cu o precizie destul de mare. Nu vedea punctul în mod conștient, dar informația vizuală ajungea în creierul lui și îi ghida mâna. Putea și să recunoască trăsături vizuale de bază, făcând, de exemplu, diferența între o linie verticală și o linie orizontală, fără să aibă conștiință vizuală. Spunea că nu vede nimic, dar pur și simplu *știa* ce era în fața lui. Avea ceea ce se cheamă o certitudine cognitivă. Acest fenomen ciudat se numește

[0] A.M. Turing, „On Computable Numbers, With an Application to the Entscheidungsproblem”, Proceedings of the London Mathematical Society S2 – 42 (1937): 230 – 265.

[0] C.E. Shannon, „A Mathematical Theory of Communication”, Bell Sistem Tehnical Journal 27 (1948): 379 – 423 și 623 – 656.

[0] R.W. Kentridge, C.A. Heywood și L. Weiskrantz, „Attention Without Awareness în Blindsight”, Proceedings: Biological Sciences 266 (1999): 1805 – 1811; R.W. Kentridge, C.A. Heywood și L. Weiskrantz, „Spațial Attention Speeds Discrimination Without Awareness în Blindsight”, Neuropsychologia 42 (2004): 831 – 835.

orbire corticală și este de obicei consecința leziunii cortexului vizual primar.

Kentridge și colegii săi au studiat pacientul GY pentru a afla ce putea și ce nu putea să vadă în câmpul „orb”. Un punct luminos se aprindea undeva în câmpul orb. Imediat apoi era afișată o linie scurtă în același loc. GY trebuia să spună dacă linia era orizontală sau verticală. Timpul lui de reacție era scurt în acea situație pentru că, aparent, punctul inițial îi atrăgea atenția și pregătea procesarea liniei. Dar, când cercetătorii afișau punctul luminos într-un loc și linia în altul, timpul lui de reacție era mai lung. Se pare că punctul îi atrăgea atenția în locul greșit. Atenția lui trebuia să își schimbe locul, ceea ce îi întârzia ușor răspunsul.

Acest experiment a constituit un moment hotărâtor în studiul conștiinței. A arătat în sfârșit că mecanismul atenției poate exista chiar și dacă mecanismul conștiinței e defect. Evident, atenția nu este doar concentrarea locală a conștiinței. E o proprietate diferită.

În ultimii douăzeci și cinci de ani, numeroase studii au confirmat faptul că conștiința poate fi separată de atenție nu doar în cazul oamenilor cu leziuni cerebrale, ci și la cei sănătoși.[0] Oamenii pot să acorde o atenție minimă unor

[0] Citez aici doar câteva dintre numeroasele cercetări excelente care demonstrează separarea dintre conștiință și atenție. Dat fiind că este unul dintre puținele rezultate legate direct de conștiință care poate fi demonstrat în laborator într-un mod controlat, acest fenomen a fost studiat îndeaproape. U. Ansorge și M. Heumann, „Shifts of Visuospatial Attention to Invisible (Metacontrast-Masked) Singletons: Clues From Reaction Times and Event-Related Potentials”, Advances în Cognitive Psychology 2 (2006): 61 – 76; P. Hsieh, J.T. Colas și N. Kanwisher, „Unconscious Pop-Out: Attentional Capture By Unseen Feature Singletons Only When Top-Down Attention is Available”, Psychological Science 22 (2011): 1220 – 1226; J. Ivanoff și R.M. Klein, „Orienting of Attention Without Awareness is Affected by Measurement-Induced Attentional Control Settings”, Journal of Vision 3 (2003): 32 – 40; Y. Jiang, P. Costello, F. Fang, M. Huang și S. He, „A Genderand Sexual Orientation-Dependent Spațial Attentional Efect of Invisible Images”, Proceedings of the National Academy of Sciences
U. S.A. 103 (2006): 17048 – 17052; R.W. Kentridge, T.C. Nijboer și

imagini vagi sau afișate într-un interval scurt, în sensul că își concentrează asupra lor resursele de procesare și chiar reacționează la ele, deși susțin că nu au văzut absolut nimic.

În absența conștiinței, atenția nu pare să funcționeze întru totul normal, pierzându-și o parte din mecanismul de control. Pentru a da un exemplu, una dintre abilitățile esențiale pe care le folosim în fiecare zi este cea de a *nu* ne concentra atenția asupra a ceva la care n-ar trebui să ne concentrăm. Lumea e plină de obiecte care țipă după atenția noastră, și uneori avem nevoie de toată forța atenției noastre ca să ne întoarcem de la un obiect (cum ar fi un țânțar care zboară prin apropiere) și să rămânem concentrați asupra altuia (cum ar fi o carte). În această situație, se poate să continuăm să fim întrucâtva atenți la țânțar și să-l monitorizăm, rămânând în același timp mai concentrați la citit. Această provocare este unul dintre cele mai evidente moduri în care schema atenției poate fi utilă. Pentru a realiza această sarcină, creierul trebuie să cunoască starea în care se află atenția în fiecare moment, să o monitorizeze și să știe când un volum prea mare de atenție trece dinspre carte spre țânțar. Trebuie de asemenea să aibă un bun model de lucru al dinamicii spațiale și temporale a acestei treceri pentru a o putea contracara. Acum să presupunem că sistemul conștiinței e defect. O parte a atenției este atrasă de țânțar, dar persoana nu e conștientă subiectiv de el. Atenția și conștiința sunt disociate. Într-un astfel de scenariu, dacă teoria schemei atenției este corectă, creierul nu știe că atenția e atrasă spre locul greșit. Modelul atenției pe care îl deține este incomplet. Oamenilor aflați în această situație

C. A. Heywood, „Attended But Unseen: Visual Attention Is Not Suficient for Visual Awareness”, Neuropsychologia 46 (2008): 864 – 869; C. Koch și N. Tsuchiya, „Attention and Consciousness: Two Distinct Brain Processes”, Trends în Cognitive Sciences 11 (2007): 16 – 22; A. Lambert, N. Naikar, K. Melachlan și V. Aitken, „A New Component of Visual Orienting: Implicit Effects of Peripheral Information and Subthreshold Cues on Covert Attention”, Journal of Experimental Psychology, Human Perception and Performance 25 (1999): 321 – 340;

le e greu să minimizeze atenția îndreptată spre țânțar. În ciuda a ceea ce e de așteptat, dacă nu ești conștient de prezența țânțarului, atenția ta se abate în mai mare măsură asupra lui și dai mai puțină atenție cărții. Experimente realizate în laboratorul meu și în alte laboratoare, care au folosit stimuli meniți să distragă atenția, asemănători țânțarului, și stimuli-țintă similari cărții, au confirmat acest tipar al rezultatelor.[0]

Se poate ca aceste numeroase experimente să dea impresia falsă că conștiința și atenția sunt ușor de separat în condiții de laborator sau că separarea se întâmplă în mod obișnuit în viața de zi cu zi. Dar să separi conștiința de atenție e ca și cum ai răzui vopseaua de pe pereți. Cele două sunt strâns legate între ele. Dacă aduci sistemul vizual la limita capacității sale în condiții controlate de laborator, cu stimuli abia perceptibili sau vagi, atunci atenția și conștiința tind să se separe. Este însă aproape imposibil să găsești un stimul vizual atât de vag, atât de scurt, atât de acoperit de alți stimuli, încât oamenii să nu fie conștienți de el, și totuși să-l pui în evidență suficient de bine încât să atragă măcar puțin din atenția cuiva. Ne-a luat câțiva ani de experimente-pilot pentru a trasa o linie între atenție și concentrare.[0] Am auzit povești asemănătoare și de la alți oameni de știință.

Conform teoriei schemei atenției, scopul conștiinței este să monitorizeze constant atenția. Astfel, conștiința urmărește îndeaproape atenția, la fel cum schema corporală urmărește poziția mâinii. Cele două se separă doar când sistemul se află în condiții de stres, când se zbate la limita capacităților sale.

---

[0] Y. Tsushima, Y. Sasaki și T. Watanabe, „Greater Disruption Duc to Failure of Inhibitory Control on an Ambiguous Distractor”, Science 314 (2006): 1786 – 1788; T.W. Webb, H.H. Kean și M.S.A. Graziano, „Effects of Awareness on the Control of Attention”, Journal of Cognitive Neuroscience 28 (2016): 842 – 851.

[0] T.W. Webb, H.H. Kean și M.S.A. Graziano, „Effects of Awareness on the Control of Attention”, Journal of Cognitive Neuroscience 28 (2016): 842 – 851.

INFORMAȚII INTEGRATE

Când iei în mână un măr, poți să procesezi culoarea, forma, mirosul, netezimea, sunetul care se aude când muști din el, gustul, reacția ta emoțională și multe alte aspecte. Fiecare dintre aceste elemente, de unul singur, poate fi procesat în creierul tău fără aportul conștiinței. Chiar și o reacție emoțională se poate desfășura în subconștient. Dar, când conștiința intră în ring, toate aceste componente sunt aduse laolaltă într-o înțelegere unitară și bogată.

Acest tip de observație a dus la perspectiva comună – una dintre puținele – potrivit căreia conștiința are de-a face cu integrarea masivă a informațiilor în creier.[0] E controversat modul anume în care interacționează cele două procese. Unele teorii susțin că conștiința unește

[0] Este imposibilă citarea tuturor teoriilor care conțin ideea potrivit căreia conștiința ține de integrarea informației. Dau aici doar câteva exemple. B.f. Baars, A Cognitive Theory of Consciousness (Cambridge, UK: Cambridge University Press, 1988); A.B. Barrett, „An Integration of Integrated Information Theory with Fundamental Physics”, Frontiers în Psychology 5 (2014): 63; F. Crick și C. Koch, „Toward a Neurobiological Theory of Consciousness”, Seminars în the Neurosciences 2 (1990): 263 – 275; A. Damasio, Self Comes to Mind: Constructing the Conscious Brain (New York: Pantheon, 2015). [V. Ed. Rom.: Șinele. Construirea creierului conștient, trad. De Doina Lică, Humanitas, București, 2016]; S. Dehaene, Consciousness and the Brain (New York: Viking Press, 2014); G.M. Edelman, J.A. Gaily și B.J. Baars, „Biology of Consciousness”, Frontiers în Psychology 2 (2012): 4; A.K. Engel și W. Singer, „Temporal Binding and the Neural Correlates of Senzory Awareness”, Trends în Cognitive Sciences 5 (2011): 16 – 25; S. Grossberg, „The Link Between Brain Learning, Attention, and Consciousness”, Consciousness and Cognition 8 (1999): 1 – 44; V.A. Lamme, „Towards a True Neural Stance on Consciousness”, Trends în Cognitive Sciences 10 (2006): 494 – 501; G. Tononi, M. Boly, M. Massimini și C. Koch, „Integrated Information Theory: From Consciousness to its Physical Substrate”, Nature Reviews Neuroscience 17 (2016): 450 – 461; C. Von der Malsburg, „The Coherence Definition of Consciousness”, în M. Ito, Y. Miyashita și E. Rolls (coord.), Cognition, Computation, and Consciousness (Oxford, UK: Oxford University Press, 1997, pp. 193 – 204); L.M. Ward, „The Thalamic Dynamic Core Theory of Conscious Experience”, Consciousness and Cognition 20 (2011): 464 – 486.

informațiile. Altele sugerează că procesul cauzal este invers: integrarea informațiilor în rețele complexe determină apariția conștiinței. Cea mai cunoscută versiune este teoria informației integrate a lui Giulio Tononi.[0] Susținătorii acestei teorii calculează un număr, phi (Φ), care reprezintă volumul de informație integrată prezent în creier, în telefonul mobil sau în orice altceva. Pe măsură ce phi crește, crește și nivelul de conștiință.

Unele detalii ale acestor teorii s-ar putea să nu fie compatibile cu teoria schemei atenției, dar toate împărtășesc acceptarea faptului că conștiința are o relație cu integrarea informației. Aici vreau să mă concentrez asupra unui punct fundamental referitor la integrare, ușor de trecut cu vederea. Într-un anumit sens, informația poate fi adezivă, și unele tipuri de informații se lipesc mai bine decât altele.

Să ne imaginăm mai multe puncte împrăștiate pe o bucată de hârtie, majoritatea negre, câteva roșii. Cele roșii alcătuiesc o formă mai mare, litera X. Forma îți sare în ochi. Informația comună despre culoare face legătura între puncte. Importanța acestei observații a fost înțeleasă pentru prima dată de psihologii gestaltiști de la începutul secolului XX, care au studiat regulile ascunse care grupează imaginile vizuale în ansambluri mai mari.[0] Nu vreau să spun că informația despre culoare este la propriu adezivă sau că atomi de informație se leagă spontan și formează molecule de informație. Dar în cazul specific al sistemului informatic cerebral, într-un mod specializat, de nivel superior, informația despre culoare poate face legătura între piese diferite pentru a construi informații compacte. Culoarea este un conector standard de informații.

Culoarea este limitată la domeniul vizual. Ne ajută să grupăm obiectele din câmpul vizual, dar nu poate conecta informații multisenzoriale. Localizarea spațială este însă un

---

[0] G. Tononi, Phi: A Voyage From The Brain To The Soul (New York: Pantheon, 2012).

[0] K. Koffka, Principles of Gestalt Psychology (New York: Harcourt, Brace and Company, 1935).

conector mai general. Dacă vezi o pasăre fremătând pe o creangă și auzi un ciripit care vine din același loc, vei avea tendința de a asocia cei doi stimuli și de a-i trata ca un obiect unic.[0]38 Făcând legătura între informația vizuală și informația auditivă, pe de o parte, și informațiile despre un loc din spațiu, pe de altă parte, creierul tău poate alătura întreaga serie într-un pachet mai mare, unitar, ca și cum ar pune laolaltă piese de puzzle.

Un motiv pentru care localizarea este un conector atât de versatil se referă la caracterul ei relațional. Informațiile despre așezarea în spațiu nu relevă o trăsătură specifică, intrinsecă, a păsării, ci descriu relația dintre tine și pasăre (e la șase metri în stânga ta), iar această relație se aplică la fel tuturor trăsăturilor specifice păsării. Culoarea, forma și sunetele, toate au *aceeași* poziție. Informațiile referitoare la localizare sunt atât de utile pentru integrarea altor fragmente de informație, încât au statutul cerebral special de facilitatori. Dacă informațiile spațiale ar fi cumva eliminate brusc din creier, lumea noastră perceptuală s-ar prăbuși într-un haos de caracteristici și simțuri separate.[0] Informațiile despre localizare sunt pentru sistemul perceptiv ca aluatul care ține crema celorlalte informații. Ar putea pe bună dreptate să se formuleze o teorie a „informației integrate” a localizării – nu pentru că informația integrată ar genera așezarea în spațiu (ceea ce nu are sens), ci pentru că, dat fiind modul în care creierul procesează lumea, informațiile despre așezarea în spațiu sunt deosebit de adezive.

Chiar dacă informațiile despre așezarea în spațiu pot face legătura între date vizuale, auditive, tactile și poate chiar olfactive, dacă ești un câine cu un foarte bun simț al mirosului, localizarea rămâne totuși un conector limitat. E

[0] B.E. Stein și M.A. Meredith, The Merging of the Senses (Cambridge, MA: MIT Press, 1993).

[0] Sindromul Bălint, cauzat de leziuni la nivelul lobului parietal, poate fi un exemplu de dezintegrare a unei lumi senzoriale unitare atunci când informația spațială este compromisă. H. Udesen și A.L. Madsen, „Balint’s Syndrome – Visual Disonentation”, Ugeskriftfor Larger 154 (1992): 1492 – 1494.

evident că unele domenii informaționale nu au o componentă spațială. Emoțiile, gândurile, convingerile, imaginația, intuițiile matematice – niciuna dintre acestea nu e ancorată într-un loc anume în spațiul din jurul corpului tău. Informațiile despre spațiu nu pot fi utilizate drept conector *universal*.

Putem să găsim un tip de informație atât de adezivă încât să funcționeze ca adeziv universal, ținând laolaltă orice tipuri de informație?

Schema atenției – informația care descrie starea atenției tale – poate fi un astfel de conector universal. Este relevantă pentru pasărea la care te uiți, pentru sunetele pe care le auzi, pentru gândurile asupra cărora te apleci, pentru emoțiile pe care le simți. Conform acestei teorii, schema atenției descrie relația generală dintre tine și un obiect. În acest sens, atenția seamănă cu așezarea în spațiu, care este tot o relație dintre tine și un obiect. Dar, spre deosebire de aceasta, relația atențională este cu adevărat aplicabilă universal. Orice e concret sau abstract, perceptual sau intelectual poate fi obiect al atenției.

Să ne imaginăm că pasărea e încă pe o creangă la stânga ta. Să analizăm acest moment specific în care ești atent la pasăre. Creierul tău construiește informații despre imaginea păsării și despre sunetele pe care le scoate aceasta, despre reacția emoțională pe care ți-o declanșează ea, emite poate chiar și o idee intelectuală despre păsări. Te ocupi de componentele A, B, C și așa mai departe. Poate ești atras în primul rând de cântecul ei și în mai mică măsură de penaj, și poate chiar ești distras de o mâncărime de pe mână care n-are nicio legătură cu pasărea, în următorul moment, atenția îți poate fi atrasă de altceva, dar în clipa de față întreaga ta atenție este îndreptată într-un mod specific spre aceste componente ale lumii. Pentru a reprezenta starea lumii tale, pe tine și relația dintre tine și lume, creierul tău trebuie să construiască un model intern de tipul A, B, C *și* un model al atenției tale și să facă legătură între toate acestea. Tratând atenția ca pe o însușire relațională a lumii pentru care

merită creat un model, creierul construiește un conector central, schema atenției, de care se vor atașa în mod necesar toate celelalte informații din câmpul atenției tale.

Accesând această rețea interconectată de informații, cogniția ta află următoarele: „Acolo e o pasăre, e colorată și cântă frumos, sunt fericit, să caut mai târziu și alte specii, mă mănâncă mâna – și toate aceste componente nu sunt separate, ci sunt aduse laolaltă sub o singură umbrelă, pentru că am o experiență subiectivă – o *conștiință* – care le asimilează pe toate în acest moment”.

Psihologii gestaltiști s-au ocupat cel mai mult de domeniul senzorial – de studierea culorii, formei, localizării, sunetului și de alte aspecte senzoriale –, încercând să înțeleagă modul în care lumea perceptuală este înțeleasă unitar. Teoria schemei atenției extinde gestaltismul prin adăugarea conectorului ultim. Conștiința unifică diverse trăsături într-un singur întreg – eu, parte din lume, în momentul prezent. Dacă informația despre așezarea în spațiu este o componentă de legătură, atunci ea se aseamănă adezivului pentru lemn, care funcționează parțial. Conștiința e mai degrabă ca un lipici universal care face legătura între toate domeniile de informație, mereu activă, indiferent ce se află în câmpul atenției noastre. Fără ea, întreaga noastră lume s-ar prăbuși într-un haos de componente care plutesc separat.

Așadar, teoria schemei atenției propune propria versiune de teorie a informației integrate. Conform acestei versiuni, nu informația integrată în sine determină în vreun fel direct apariția conștiinței; ci informația din creier are capacitatea de a fi adezivă, unele tipuri de informații fiind mai adezive decât altele. Culoarea poate contribui la integrarea informației în câmpul vizual. Localizarea în spațiu poate contribui la integrarea unui volum mare de informații din cea mai mare parte a lumii senzoriale. Probabil că multe alte tipuri de informație sunt adezive în grade diferite. Cea mai adezivă informație, cea care are cel mai înalt nivel de conectivitate, este informația despre relația atențională dintre tine și elementele din lumea ta.

Prin definiție, fiecare element din câmpul atenției tale are această proprietate. Ca urmare a acestei proprietăți *über Gestalt*, conștiința guvernează integrarea masivă a informației în creier.

8
Mașinării conștiente

Sir Isaac Newton, marele fizician și matematician, și-a încercat norocul cu alchimia.[0] Judecând după numărul de caiete pe care le-a umplut, se pare că a petrecut mai mult timp încercând fără succes să creeze aur decât formulând, cu succes, legea gravitației. Una dintre rețetele lui includea Dragonul Arzător, Porumbeii Dianei și Vulturii lui Mercur, toate denumiri exotice pentru ingrediente reale. Îmi imaginez cum Newton și alți alchimiști se aplecau asupra pergamentelor unsuroase, amestecând și încălzind, așteptând cu nerăbdare rezultatul, doar ca să fie dezamăgiți de fiecare dată. Nu le lipseau deloc teoriile, dar își bazau cunoașterea alchimică pe analogie și mit, nu pe deducție logică. Practicienii nu puteau să spună: „Dat fiind ce știm despre proprietățile materialelor, dacă amesteci A și B, în mod logic vei obține aur”. În cel mai bun caz, alchimistul aspirant trebuia să se mulțumească cu o abordare de tip încercare și eșec.

Nu vreau să iau în râs alchimia. În încercările de atunci se află cel mai probabil originea chimiei moderne. Și, oricum, marele secret al alchimiei a fost în cele din urmă descoperit de știința modernă.

Uite rețeta, în caz că vrei s-o încerci. Iei un element mai greu decât aurul, îl zdrobești într-un accelerator de particule și în cele din urmă obții câțiva atomi de aur. Sau bombardezi un element mai ușor cu nucleoni și, într-un final, unele componente se vor lipi și vor forma un nucleu

[0] M. White, Isaac Newton: The Last Sorcerer (New York: Basic Books, 1999).

de aur. Nu știu dacă a-ncercat cineva, dar, dat fiind că teoria e corectă din punct de vedere științific, știu că funcționează. Au fost create elemente mult mai grele și mai rare decât aurul. Profitul nu e prea mare – e nevoie de milioane de dolari pentru a crea un atom de aur –, dar principiul funcționează.

Construirea unei mașinării conștiente seamănă cu alchimia. Mare parte din cercetările științifice asupra conștiinței conțin o mulțime de analogii vagi și uneori chiar mituri. Dacă luăm un nivel suficient de complexitate sau de feedback sau o seamă de conexiuni ample sau o buclă talamus-cortex impresionantă și le amestecăm pe toate deasupra unui arzător Bunsen, se va trezi mașinăria și va fi conștientă? Va afirma convingător: „Am o experiență conștientă a culorii roșii și a frigului și a propriei existențe!”

Pentru mine teoria schemei atenției este răspunsul tehnic la misterul alchimic al conștiinței. Dacă vei construi o mașinărie conform acestei teorii, adăugând modelele interne corecte și oferindu-i acces cognitiv și lingvistic la aceste modele, atunci mașinăria va avea capacitățile dorite. Nu trebuie să speri că va apărea conștiința dintr-o fuziune alchimică a ingredientelor. Mașinăria va crede că are conștiință, va afirma că are conștiință și va vorbi despre conștiința ei, întrucât o vei fi programat astfel încât să conțină constructul de conștiință.

Cursa pentru construirea de mașinării conștiente a avut un început lent. Au existat unele încercări preliminare bazate pe diverse teorii, dar nu a ieșit nimic verificabil care să aducă a conștiință.[0] Unul din lucrurile care atenuează entuziasmul este că teoriile conștiinței tind să se

---

[0] I. Aleksander, Impossible Minds: My Neurons, My Consciousness (Singapore: World Scientific, 2014); B.J. Baars și S. Franklin, „Consciousness is Computational: The LIDA Model of Global Workspace Theory”, International Journal of Machine Consciousness 1 (2009): 23 – 32; A. Chelia și R. Manzotti, „Machine Consciousness: A Manifesto for Robotics”, International Journal of Machine Consciousness 1 (2009): 33 – 51; L.A. Coward și R. Sun, „Criteria for an Efective Theory of Consciousness and Some Preliminary Attempts”, Consciousness and Cognition 13 (2004): 268 – 301;

concentreze asupra aspectelor metafizice, și nu asupra beneficiilor practice ale conștiinței. Dacă ești un specialist în tehnologia informației interesat de produse utile și vandabile, de ce ți-ai pierde timpul cu metafizica? Teoria schemei atenției oferă însă unele beneficii practice. Conform acesteia, creierul a ajuns pe parcursul evoluției să dezvolte conștiința pentru că aceasta oferă două avantaje semnificative: primul este că îmbunătățește reglarea internă, iar al doilea este că servește ca fundație a cogniției sociale.

Având o teorie specifică și beneficii practice specifice, conștiința artificială ar putea să-și ia avânt. S-ar putea să capete existență, într-o formă sau alta, în următorii zece ani. E probabil ca primele experimente legate de conștiința artificială să fie extrem de limitate. S-ar putea să avem mașinării conștiente de imagini vizuale, dar încă lipsite de conștiință în alte privințe. La crearea de androizi inteligenți ca Data din *Star Trek sau* ca C_3PO din *Războiul stelelor*, care nu sunt doar conștienți, ci și aproape omnicapabili și străluciți în conversație, se va ajunge mult mai târziu în viitor. Este posibilă existența acestora, dar încă nu avem tehnologia necesară.

Însă, înainte să perorez despre viitorul companionilor artificiali și conștienți, trebuie să pornesc de la o întrebare fundamentală. Dacă am reuși să construim o mașinărie care să creadă despre sine că e conștientă, cum vom ști că e într-adevăr conștientă? Putem verifica prezența conștiinței?

În 1950, matematicianul Alan Turing a propus o modalitate de a testa dacă o mașinărie poate să gândească.[0] Testul implică un joc pe care îl joacă trei oameni: unul minte, unul spune adevărul și unul ghicește. Cei trei se află în încăperi diferite și nu pot comunica decât în scris, pentru a fi evitată transmiterea semnalelor personale involuntare.

---

[0] A.M. Turing, „Computing Machinery and Intelligence”, Mind 59 (1950): 433 – 460.

În prima fază, cel care ghicește nu știe nimic despre ceilalți doi, în afară de faptul că unul e bărbat și celălalt este femeie. Trebuie să ghicească cine e bărbat și cine e femeie. Cel care spune adevărul trebuie să-i ofere răspunsul corect. Cel care minte trebuie să-l inducă în eroare și să-i ofere răspunsul greșit. Cei trei sunt angajați într-un complicat război informațional. Dat fiind că sunt posibile orice întrebare și orice răspuns, conversația poate ajunge la aluzii sociale și la iscusite manipulări psihologice. La sfârșitul jocului, cel care ghicește trebuie să facă o alegere. Dacă are dreptate, cel care spune adevărul câștigă. Dacă se înșală, câștigă cel care minte. Fără ca acest joc să se fi jucat în realitate – și nu cunosc pe nimeni care să-l fi jucat –, e dificil de estimat rata tipică de succes a celui care minte sau a celui care spune adevărul. Cred că cel care spune adevărul e în avantaj, dat fiind că dacă spui adevărul nu poți să cazi în capcana unei contradicții. În orice caz, dacă jocul se joacă de mai multe ori, poate fi estimată rata de succes a jucătorilor.

Următorul pas constă în înlocuirea omului care minte cu o mașinărie care minte. Dacă mașinăria câștigă jocul la fel de des ca un om, atunci mașinăria gândește ca un om. Acesta este testul Turing inițial.

La prima vedere, testul lui Turing poate părea prea complicat. Cu siguranță este nepractic, dat fiind că trebuie efectuat de multe ori pentru a se ajunge la date statistice. Dar, privit mai îndeaproape, este strălucit și avansat pentru vremea sa. Pentru a se descurca bine, mașinăria are nevoie de o teorie a minții altora. Mai mult decât atât, trebuie să treacă testul convingerilor false. Trebuie să înțeleagă că alți oameni, alte minți, pot avea convingeri greșite despre lume și trebuie să știe ce convingere aparține cui. E un test de cogniție socială foarte greu de trecut.

Testul lui Turing poate fi considerat o versiune timpurie a testului Sally-Anne (al convingerilor false), propus pentru prima dată de psihologi în anii 1980.[0] Am prezentat acest

[0] S. Baron-Cohen, A.M. Leslie și U. Frith, „Does the autistic child have a «theory of mind?»”, Cognition 21 (1985): 37 – 46; H. Wimmer și J.

test în Capitolul 5, dar îl voi rezuma și aici. În versiunea obișnuită, psihologul care îți testează cogniția socială îți va spune o poveste despre cum Sally e păcălită de Anne. Sally își pune sandvișul în coșul A și se duce la toaletă, iar cât timp e plecată, Anne mută sandvișul în coșul B. Când Sally se întoarce, în ce coș va căuta mai întâi sandvișul? Dacă înțelegi că Sally are o minte, că mintea ei are convingeri și că acestea pot fi contrare realității, atunci poți rezolva sarcina. Sally va căuta mai întâi în coșul A, unde a lăsat sandvișul. Oricât de simplă le-ar părea această sarcină adulților care au ani de experiență socială, copiii sub cinci ani nu o pot rezolva, și foarte puține animale nonumane prezintă dovezi că ar putea să o rezolve.[0]

Testul Sally-Anne, implicând modele mintale și convingeri greșite, este în fond o versiune simplificată a testului lui Turing. Se poate ca Turing să se fi gândit că a conceput un test de tehnologie a informației, dar e evident că era un bun psiholog social și că era cu cel puțin treizeci de ani înaintea timpului său. Pentru a verifica dacă o mașinărie poate să gândească, aș prefera testul complicat al lui Turing în locul versiunii simplificate Sally-Anne. Mașinăria s-ar confrunta cu o problemă mai complicată. Pentru a se descurca la fel de bine ca un om, ar trebui să aibă abilități spectaculoase de limbaj, o bună cunoaștere a vieții de zi cu zi și o capacitate extraordinară de a face raționamente despre convingerile altor oameni.

Turing nu a afirmat niciodată că jocul său ar fi un test de conștiință. Nu a discutat relația acestuia cu experiența subiectivă, conștiința, qualia – senzațiile interioare care

---

Ferner, „Beliefs About Beliefs: Representation and Construining Function of Wrong Beliefs în Young Children's Understanding of Deception", Cognition 13 (1983): 103 – 128.

[0] N.S. Clayton, „Ways of Thinking: From Crows to Children and Back Again", Quarterly Journal of Experimental Psychology 68 (2015): 209 – 241; C. Krupenye, F. Kano, S. Hirata, J. Call și M. Tomasello, „Great Apes Anticipate That Other Individuals Will Act According to False Beliefs", Science 354 (2016): 110 – 114; H.M. Wellman, D. Cross și J. Watson, „Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief", Child Development 72 (2001): 655 – 684.

apar odată cu procesarea informației – sau cu orice alt fel de noțiune asociată cu conștiința. Dacă o mașinărie trece testul lui Turing în forma sa inițială, atunci putem fi siguri că e un computer extrem de sofisticat. Trebuie să aibă ceva care seamănă cu *conținutul* normal al conștiinței umane, altfel n-ar putea să țină pasul în conversație; dar nu există nicio garanție că are experiență conștientă.

În anii care au urmat primei sale publicări de către Turing însuși, testul a fost preluat și modificat de un club enorm de fani din domeniul inteligenței artificiale. Celebrul test Turing de care majoritatea oamenilor au auzit astăzi diferă destul de mult de original și de testul Sally-Anne. Versiunea modernă pune accentul pe conștiință, nu pe cogniția socială. Pentru a-ți da seama dacă o mașinărie e conștientă, nu trebuie decât să ai o conversație cu ea. Atunci când nu-ți poți da seama dacă e vorba de o mașinărie sau de o persoană, mașinăria trece testul. Noul test este mult mai simplu decât originalul și poate fi realizat cu ușurință în viața reală, ceea ce s-a și întâmplat de multe ori. Au fost organizate conferințe întregi în care astfel de mașinării au fost supuse testului Turing modernizat, cea mai celebră fiind competiția anuală Premiul Loebner, deși până acum nicio mașinărie nu a reușit să-i convingă pe oamenii de știință că ar fi conștientă.

Chiar dacă o mașinărie ar putea trece versiunea modernă a testului Turing, tot n-am ști dacă are experiență interioară. Un test de abilități conversaționale este un test ilogic de conștiință. Dat fiind cât de ușor sunt de convins unii oameni că râurile și copacii au conștiință, s-ar putea să nu fie prea dificil pentru aceștia să creadă că o mașinărie e conștientă, deși nu e. Dacă am putea construi o mașinărie cu adevărat conștientă, s-ar putea ca aceasta să pice testul Turing. Conștiința nu vine neapărat la pachet cu capacitatea de a conversa sofisticat și politicos. La naiba, un copil de trei ani n-ar putea trece testul Turing. Nici unii adulți. Niciun câine, deși majoritatea oamenilor sunt destul de siguri că și câinii au experiență conștientă. Pur și simplu nu e un test care să facă o diferență clară. Dar ideea

testului Turing a ajuns atât de familiară, atât de răspândită cultural, încât se pare că oamenii au acceptat-o ca soluție practică. Raționamentul se desfășoară cam așa:

Știu că sunt conștient pentru că am o experiență directă a propriei minți. Dar nu pot să știu cu adevărat dacă alți oameni sunt conștienți. Oricât mi-aș dori să cred asta, oricât mi-aș iubi copiii, și soția, și pisica, nu pot avea experiența directă a conștiinței lor. E imposibil. Nu pot decât să aplic cea mai bună soluție și să *presupun* provizoriu că sunt conștienți pentru că se poartă ca și cum ar fi. În mod similar, nu pot să dovedesc existența conștiinței în cazul unei mașinării. Această ipoteză este nedemonstrabilă. Dar testul Turing e cea mai bună soluție. Pot să verific dacă mașinăria face parte din aceeași categorie ca soția mea, copiii mei și animalele mele de companie. Dacă se poartă la fel de convingător ca o ființă conștientă, pot să presupun provizoriu că e conștientă.

Cam așa sună raționamentul obișnuit. Dar vreau să deconstruiesc mistica testului Turing. Din perspectiva teoriei schemei atenției, *putem* să știm cu certitudine obiectivă dacă o mașinărie are același tip de conștiință ca oamenii. Iar experiența personală directă *nu* e unicul mod – nici măcar nu e un mod bun – de a ști dacă cineva e conștient.

„*Bineînțeles* că sunt conștient. Știu că sunt, pentru că am experiența directă a acestui lucru”.

Iată un raționament ciclic în toată splendoarea lui. Conștiința *este* o experiență directă. Așadar, afirmația este echivalentă cu: „Știu că sunt conștient pentru că sunt conștient”. Așa cum am mai spus, mașinăria este captivă informației pe care o conține. Un anumit model intern informează mașinăria că are conștiință și că, așadar, „știe” că e conștientă. Modelul intern informează mașinăria că conștiința ei nu are substanță fizică și rămâne cu desăvârșire o experiență subiectivă, așadar „știe” că conștiința ei nu poate fi confirmată de nimeni altcineva.

Dar un model intern înseamnă informație, iar informația

poate fi măsurată obiectiv. Nu trebuie să ne bazăm pe afirmații subiective. Conform teoriei schemei atenției, pentru a determina dacă o mașinărie e conștientă, trebuie să-i verificăm mecanismele, pentru a descoperi care dintre ele conține o schemă a atenției, și trebuie să vedem ce informații sunt incluse în schema atenției. Vom afla astfel cu certitudine dacă avem de-a face cu o mașinărie care crede că are experiență conștientă subiectivă în același mod în care credem noi că avem. Dacă modelul său intern conține informațiile necesare, atunci da. Dacă nu, nu. Toate acestea pot fi, în principiu, măsurate și confirmate.

Măsurarea informației din creierul uman este dificilă, dar nu imposibilă din punct de vedere practic. Oamenii de știință o practică deja într-un mod limitat. Dacă îți atașează o serie de electrozi pe cap, oamenii de știință pot să vadă dacă ai de gând să-ți miști mâna la dreapta sau la stânga.[0] Cu un RMN de înaltă calitate al cortexului vizual, pot să vadă dacă te uiți la o față sau la o casă.[0] Citirea informației mai complexe nu e posibilă încă, dar ar trebui să fie o chestiune de progres tehnologic. Sunt sigur că, în timp, va apărea tehnologia necesară descifrării informațiilor din creier la o rezoluție înaltă (un gând foarte tulburător). În principiu, ar trebui să fie posibilă măsurarea informației din schema atenției, chiar dacă acest tip de extragere a informației sofisticate din creier ține încă de viitorul îndepărtat. Ce vreau să spun este că, potrivit teoriei schemei atenției, conștiința nu e, prin definiție, indiscutabil intimă. Nu e adevărată afirmația: „Știu că eu sunt conștient, dar nu am cum să știu dacă și tu ești”. A afla dacă un creier crede că are o conștiință similară cu a mea ține de inventarea mijloacelor tehnologice necesare citirii informației prezente în acel creier.

---

[0] M.A. Lebedev și M.A. Nicolelis, „Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation”, Physiological Review 97 (2017): 767 – 837.

[0] J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Schouten și P. Pietrini, „Distributed and Overlapping Representations of Faces and Objects în Ventral Temporal Cortex”, Science 293 (2001): 2425 – 2430.

În cazul inteligenței artificiale, dat fiind că nu există un creier biologic, ar trebui să fie mult mai simplu să te uiți sub capotă și să măsori conștiința. Ar trebui ca tehnicienii să poată să desfacă o mașinărie construită de om și să măsoare conținutul informațional, dat fiind că avem instrumente mult mai potrivite pentru acest tip de demers decât avem pentru creier, și pentru că probabil cineva are planurile după care mașinăria a fost construită.

Confuzia apare când oamenii spun: „Înțeleg că poți să măsori informația, dar cum știi dacă mașinăria are și senzații? Doar mașinăria poate avea direct experiența propriei conștiințe, așadar cum poate ști altcineva ce simte?” Această întrebare apare când cogniția ta consultă schema atenției. Această schemă a atenției îți spune că ai o senzație intimă, personală, imaterială, la care doar tu ai acces. Dar mașinăria noastră ipotetică are toate aceste componente pe care le ai și tu, inclusiv o schemă a atenției care conține aceeași informație. Consultând schema atenției, și ea va fi informată că are o senzație intimă, personală și imaterială. Și tu, și mașinăria, fiind construiți la fel, sunteți blocați în același raționament circular. Nu știi decât ce știi și nu poți decât să te raportezi la informațiile pe care le deții – iar aceste informații pot fi, în principiu, evaluate de altcineva.

Ce vreau să spun este că întrebarea aparent fără răspuns – cum am putea ști dacă o mașinărie e conștientă? — poate, în principiu, să primească un răspuns definitiv odată ce avem echipamentele adecvate pentru evaluarea informației. Nu trebuie să ne bazăm pe testul Turing, care e indirect, în cel mai bun caz, și, în cel mai rău caz, irelevant în privința conștiinței.

În continuare, voi discuta utilitatea construirii unei mașinării conștiente folosind ca ghid teoria schemei atenției. Avem nevoie de patru componente. Prima: mașinăria trebuie să aibă atenție artificială – capacitatea de a-și concentra resursele și de-a le îndrepta asupra unui obiect și apoi asupra altuia. A doua: mașinăria trebuie să

aibă o schemă a atenției – un model intern care descrie atenția într-un mod general și, făcând asta, informează mașinăria cu privire la conștiința subiectivă. A treia: mașinăria trebuie să aibă o gamă adecvată de conținuturi. De exemplu, putem să avem o mașinărie care nu are decât conștiință vizuală – dar, cu o gamă atât de limitată, nu prea ar aduce a om. În mod ideal, ar avea conținuturi mult mai ample. A patra: mașinăria trebuie să aibă un motor de căutare sofisticat care poate accesa modele interne și poate vorbi despre acestea, astfel încât să putem purta o conversație cu ea și să ne poată vorbi despre conștiința ei. Dacă putem construi aceste patru componente, obținem o mașinărie similară întrucâtva conștiinței umane.

Să analizăm fiecare componentă în parte și să vedem cât e de practică.

Prima componentă e atenția. Mașinăria ar trebui să aibă capacitatea de a-și concentra resursele de procesare și de a-și schimba focalizarea de la obiect la obiect – de la un măr din apropiere la o gogoașă, la o persoană din cealaltă parte a încăperii, la un eveniment întru totul intern, cum ar fi o amintire.

Atenția, sub o formă sau alta, există deja la dispozitivele artificiale. Gestionarea sistematică a resurselor este ceva obișnuit în orice computer modern. În plus, mulți cercetători au schițat sau au construit modele computaționale – simulări computerizate – care ilustrează diferite aspecte ale atenției umane.[0]

---

[0] Au fost construite atât de multe dispozitive artificiale dotate cu atenție, încât nu pot oferi aici decât o listă incompletă. H. Adeli, F. Vitu și G.F. Zelinsky, „A Model of the Superior Colliculus Predicts Fixation Locations during Scene Viewing and Visual Search”, Journal of Neuroscience 37 (2017): 1453 – 1467; A. Borji și L. Itti, State-of-the-Art în Visual Attention Modeling”, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2013): 185 – 207; G. Deco și E.T. Rolls, „A Neurodynamical Cortical Model of Visual Attention and Invariant Obiect Recognition”, Vision Research 44 (2004): 621 – 642; Y. Fang, C. Zhang, J. Li, J. Lei, M. Perreira da Silva și P. Le Callet, „Visual Attention Modeling for Stereoscopic Video: A Benchmark and Computational Model”, IEEE

Dar pentru construirea unei mașinării conștiente e nevoie de un tip anume de atenție, chestiune nerezolvată în mod satisfăcător. Atenția umană are o gamă specifică de proprietăți. Când e concentrată asupra unui lucru, îți permite să procesezi acel lucru în detaliu și să extragi din el o semnificație profundă. Atenția îți permite să identifici disponibilitățile lucrului – poate fi apucat, lovit, mușcat? Atenția te face să poți acționa asupra lui – îți permite să iei decizii cu privire la ce să faci cu el. Îți permite să-l stochezi în memorie pentru mai târziu.

Dinamica atenției te face să focalizezi rapid procesarea pe stimulii puternici sau care ies în evidență, dar, în același timp, nu este complet aservită mediului – o poți controla, comenzi interne îndreptând-o spre un obiect sau altul. Și mai important este faptul că atenția poate trece cu ușurință de la un domeniu informațional la altul. Poate fi direcționată spre puncte spațiale (principala proprietate a atenției care a fost încorporată până acum în sisteme artificiale), dar poate fi focalizată și pe culoare, mișcare, gust, presiune tactilă sau pe ceva intern, cum ar fi o amintire sau o idee.

Nu cunosc până acum în lumea inteligenței artificiale exemple de atenție de un asemenea grad de complexitate. Un motiv este că inteligența artificială tinde să fie dependentă de domeniu. De exemplu, să luăm cazul recunoașterii faciale. E o funcție importantă cu diverse aplicații, dar în principiu nu include chestiunea atenției. O versiune simplă de atenție spațială poate să îndrepte camera spre o față pe care vrei să o identifici și algoritmul de recunoaștere intră în funcțiune. Mașinăria nu trebuie să-și schimbe ținta atenției de la un domeniu de informație la altul, de la fețe la sunete, la gânduri, pentru că nu funcționează decât în domeniul recunoașterii faciale. În esență, sistemul nu are nicio relație cu atenția în afară de sensul cel mai simplu și mai evident de a-și îndrepta ochii

---

Transactions on Image Processing 26 (2017): 4684 – 4696; S. Goferman, L. Zelnikmanor și A. Tai, „Context-A ware Saliency Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012): 1915 – 1926; C. Guo și L. Zhan

spre o țintă. Atenția artificială este posibilă din punct de vedere tehnic, și sunt mulți oameni care lucrează la asta, dar până acum nu are generalitatea atenției umane. Se fac progrese însemnate în privința acestei prime componente a unei mașinării conștiente, dar mai e cale lungă până la un rezultat.

A doua componentă de care mașinăria are nevoie e o schemă a atenției, modelul intern esențial care descrie atenția într-o manieră generală și, prin asta, informează mașinăria cu privire la conștiință.

Un coleg mi-a spus odată că orice computer are un fel de schemă a atenției, o serie de informații care monitorizează distribuirea resurselor de procesare. Astfel, conform teoriei mele, orice computer e deja conștient. Un alt coleg mi-a spus că i-ar fi ușor să programeze întreaga teorie într-o jumătate de zi. Ar închega o versiune artificială de atenție vizuală, i-ar atașa un modul care urmărește variabilele, și gata, dispozitivul ar fi conștient, dacă ceea ce susțin eu ar fi corect. Un grup de colegi chiar a construit un dispozitiv cu atenție artificială care avea o schemă a atenției pentru monitorizarea și controlul atenției.[0] Modelul intern chiar făcea mașinăria mai eficientă în privința controlului atenției, după cum era de așteptat. Dar era aceasta cu adevărat conștientă?

Sunt destul de sigur că răspunsul e nu. Niciunul dintre aceste exemple nu posedă conștiință. O schemă a atenției nu e un talisman magic care determină apariția conștiinței dacă e plasat într-un computer. E un pachet de informații. Dacă acesta ar face o afirmație simplă, precum „Variabila A se încadrează în nivelurile acceptabile de toleranță”, și dacă mașina ar putea să verbalizeze această informație, nu

[0] E. Van den Boogaard, J. Treur și M. Turpijn, „A Neurologically Inspired Neural Network Model for Graziano’s Attention Schema Theory for Consciousness”, International Work Conference on the Interplay Between Natural and Artificial Computation: Natural and Artificial Computation for Biomedicine and Neuroscience, partea I (2017): 10 – 21.

ți-ar spune: „Am o experiență conștientă!”; ți-ar vorbi despre variabila A. Nu e nimic magic în relația dintre o schemă a atenției și conștiință. Dacă schema atenției i-ar transmite mașinăriei că are conștiință, atunci mașinăria ar fi informată că are conștiință.

Schema umană a atenției, fiind produsul a sute de milioane de ani de evoluție, are un conținut biologic ciudat și neuniform. Prezintă atenția ca proprietate invizibilă, ca pe o minte care poate să experimenteze sau să asimileze lucruri, o forță care îmi permite să acționez și să-mi amintesc, ceva care în sine nu are substanță fizică, dar care se află undeva înăuntrul meu. Schema atenției înseamnă mai mult decât un indicator care arată spre obiecte sau decât o serie de linii de cod. Construiește o imagine bogată a atenției și a consecințelor ei previzibile. Dacă s-ar construi o mașinărie cu *acest* tip de schemă a atenției, care conține aceste pachete biologice de informații neuniforme, ar rezulta o mașinărie care ar putea să afirme că e conștientă la fel ca oamenii. Niciun obstacol tehnologic major nu stă în calea unei scheme artificiale a atenției. Poate că e cea mai ușor de construit și cu siguranță e cea mai specifică dintre componentele mașinăriei.

Pasul următor este că mașinăria are nevoie de conținut. Construirea de conținut bogat și variat s-ar putea să fie partea cea mai dificilă a provocării, dat fiind că e o problemă deschisă. Paradoxal, așa-numita problemă dificilă (să faci mașinăria să fie conștientă de ceva) s-ar putea să fie partea ușoară, iar problema simplă (să-i oferi mașinăriei o gamă de materiale de care să fie conștientă) s-ar putea să fie partea dificilă. Eu unul cred că eforturile de construire a conștiinței artificiale vor începe de la imputul senzorial, în special de la văz, dat fiind cât de multe se cunosc deja despre funcționarea cerebrală a sistemelor senzoriale și despre modul în care interacționează cu atenția. Dar cred că toți am fi puțin sceptici cu privire la o mașinărie care e conștientă de un punct negru pe ecran, și de nimic altceva. Ar putea fi un test util de laborator, dar nu am considera

mașinăria pe deplin conștientă.

Chiar dacă i-am oferi mașinăriei o bogată conștiință senzorială, în cele din urmă ar trebui să poată trece dincolo de simțuri și să încorporeze în conștiință gânduri abstracte. Aici devine problema tehnică deosebit de dificilă. Nu se știu multe despre modul în care gândirea abstractă se intersectează cu mecanismele atenției și ale înțelegerii. Știm că o persoană poate fi atentă la un gând, păstrându-l în minte și concentrându-și resursele asupra acestuia în defavoarea altor gânduri posibile sau chiar a procesării senzoriale.[0] Tuturor ni s-a întâmplat să fim atât de pierduți în gânduri, încât să nu mai știm ce se întâmplă în jurul nostru. Dar neuroștiințele nu au descoperit încă ce presupune din punct de vedere mecanic acest nivel cognitiv al atenției. Vor mai trece probabil zeci de ani până să fie construită o mașinărie cu aceste proprietăți.

Dar presupun că cel mai greu de rezolvat este misterul emoției. E cel mai puțin cunoscut domeniu informațional din creier. Tot ce se poate spune în acest sens e că au fost localizate numeroase structuri cerebrale responsabile de emoție.[0] Sunt numite uneori rețele afective sau circuite afective, dar sistemul nu e înțeles la nivelul de precizie al unui circuit.

Una dintre regiunile cerebrale relevante este hipotalamusul, o mică structură, cam de dimensiunea unei nuci, situată la baza creierului. Acum mai bine de o sută de ani, fiziologul elvețian Walter Hess a descoperit că, dacă monta un electrod în hipotalamus și îi transmitea o undă ușoară de curent electric pentru a activa neuronii, putea induce reacții care păreau stări afective.[0] Cercetarea lui a fost confirmată de multe ori de atunci[0]. Frica, furia, dorința

---

[0] M.M. Chun, J.D. Golomb și N.B. Turk-Browne, „A Taxonomy of External and Internal Attention”, Annual Review of Psychology 62 (2011): 73 – 101.

[0] J. Ledoux, The Emotional Brain: The Mysterious Underpinings of Emoțional Life (New York: Simon and Schuster, 1998).

[0] W.R. Hess, Funcțional organization of the diencephalons (New York: Grune and Stratton, 1957).

[0] B.G. Hoebel, „Neuroscience and Appetitive Behavior Research: 25

sexuală, foamea – toate aceste stări pot fi evocate prin stimularea unor părți diferite ale hipotalamusului.

Un alt procesor central al emoției este amigdala, o structură cam de dimensiunea unei migdale (denumirea provine de la cuvântul „migdală” din limba greacă). Avem câte una în fiecare parte a creierului și se activează în timpul experiențelor afective, în special când asociem o emoție anume cu o situație sau o imagine[0]. Dacă te enervezi atunci când vezi un politician pe care îl detești, amigdala ta face asociații între imagine și emoție. O altă regiune cerebrală implicată profund în emoție este partea cea mai de jos a cortexului prefrontal, aflată chiar în spatele cavității osoase a ochilor. Se pare că joacă un rol în luarea deciziilor în funcție de conținutul emoțional[0]. Deși toate aceste date neuroștiințifice sunt fascinante, nu ne spun prea multe despre cum să construim emoții artificiale sau despre cum să le încorporăm în mecanismele atenției și ale conștiinței.

Un indiciu ar putea veni de la o experiență comună tuturor oamenilor – emoțiile separate de conștiință[0]. În mod normal, nu ne gândim că emoțiile au o existență separată de conștiință, dar acest lucru se poate întâmpla. Ca exemplu din viața de zi cu zi, se poate să te simți ușor

---

Years”, Appetite 29 (1997): 119 – 133; T.V. Sewards și M.A. Sewards, „Representations of Motivațional Drives în Mesial Cortex, Medial Thalamus, Hypothalamus and Midbrain”, Brain Research Bulletin 61 (2003): 25 – 49; A. Venkatraman, B.L. Edlow și M.H. Immordino-Yang, „The Brainstem în Emotion: A Review”, Frontiers în Neuroanatomy 11 (2017): 15.

[0] J. Ledoux, „The Amygdala”, Current Biology 17 (2007): R868-R874; P.f. Walen și E.A. Phelps, The Human Amygdala (New York: Guilford Press, 2009).

[0] E.T. Rolls și F. Grabenhorst, „The Orbito-frontal Cortex and Beyond: From Affect to Decision-Making”, Progress în Neurobiology 86 (2008): 216 – 244.

[0] M. Tamietto și B. De Gelder, „Neural Bases of the Non-Conscious Perception of Emoțional Signals”, Nature Reviews Neuroscience 11 (2010): 697 – 709; P. Winkielman și K.C. Berridge, „Unconscious Emotion”, Current Directions în Psychological Science 13 (2004): 120 – 123.

agitat, stresat sau furios, și să nu știi exact de ce. Apoi un prieten îți va spune: „Sigur nu te-apasă nimic?”, iar tu îți vei redirecționa atenția – și conștiința – și îți vei da brusc seama: „Ah, chiar că *sunt* stresat”. Unii oameni sunt mai conștienți de propriile emoții, alții mai puțin, dar toți avem momente în care generăm emoții fără să știm că facem asta. Orice teorie a experienței emoționale trebuie să ia în calcul această delimitare dintre stările afective și experiența conștientă.

Specialistul în neuroștiințe Joseph Ledoux a formulat o expunere elegantă a conștiinței afective pornind de la cercetările sale de pionierat în domeniul sistemelor afective cerebrale[0]. Stările afective sunt organizate în afara conștiinței de către structuri cerebrale profunde, situate sub cortex – cum ar fi hipotalamusul și amigdala, despre care tocmai am vorbit. Pentru a deveni conștient de o stare afectivă e necesar ca informația din aceste structuri profunde să ajungă la sistemul cortical, unde e integrată informațiilor cognitive despre conștiință. Conform ipotezei lui, conștiința afectivă conține două părți: informația care definește starea emoțională și informația care definește conștiința. Așa cum creierul poate opera cu seria compusă de informații: „Sunt conștient de măr”, la fel poate opera cu: „Sunt conștient că am această emoție”.

Chiar și având o astfel de cunoaștere, în practică tehnicianul nu va putea răspunde la întrebarea: „Ce informație definește emoția?” Răspunsul e că de fapt nimeni nu știe.

Un indiciu vine de la celebra teorie a emoției propusă de psihologii William James și Carl Lange în secolul al XIX-lea.[0] Conform teoriei James-Lange, emoția începe de la o senzație corporală. Inima bate mai tare, stomacul secretă

---

[0] J.E. Ledoux și R. Brown, „A Higher-Order Theory of Emoțional Consciousness”, Proceedings of the National Academy of Sciences, U.S.A. 114 (2017): E2016 Nature Reviews, Neuroscience $E_2$025.

[0] W. Cannon, „The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory”, The American Journal of Psychology 39 (1927): 106 – 124.

acid, pielea se răcește și se umezește. Apoi, detectând aceste modificări corporale și evaluând contextul, creierul construiește o poveste: „Sunt anxios” sau „Sunt tensionat”. Una dintre cele mai bune demonstrații ale acestui efect a fost celebrul experiment de pe podul din Vancouver realizat în anii 1970.[0] Experimentul presupunea ca o femeie să oprească un bărbat care mergea pe stradă și să-l roage să răspundă la câteva întrebări pentru un chestionar. În unele cazuri, bărbații erau opriți într-o situație generatoare de anxietate, în mijlocul unui pod care se legăna deasupra unui hău. În alte cazuri, bărbații erau opriți pe un pod solid, stabil. La final, bărbații erau întrebați cât de atrăgătoare li se păruse femeia care le luase interviul. Bărbații de pe podul instabil au evaluat femeia ca fiind mai atrăgătoare decât bărbații de pe podul stabil. Probabil că legănarea podului a determinat o accelerare a ritmului cardiac și le-a făcut pielea să devină rece și umedă, modificări care au fost atribuite în mod greșit atracției sexuale.

În psihologia modernă, teoria James-Lange este considerată și astăzi în mare măsură corectă. O parte din emoție este ancorată în senzațiile fizice din corp, iar o altă parte implică o reprezentare bogată, de nivel înalt, cele două interacționând într-un mod complicat, care nu este încă bine înțeles. În măsura în care teoria James-Lange este corectă, ne face să ne punem niște întrebări interesante despre emoțiile mașinăriei dotate cu inteligență artificială. Ar avea un android nevoie de stomac, de glande sudoripare și de o inimă pentru a simți emoții așa cum simt oamenii? Dacă mașinăria nu ar simți niciodată gustul mâncării sau nu ar avea senzații digestive, dacă adrenalina nu i-ar tensiona corpul astfel încât s-o facă să lupte sau să fugă, ar putea să aibă emoții reale? Presupunerea mea e că ar putea avea procese foarte similare – deși, probabil, nu identice – emoției umane. S-ar putea să trebuiască să-i implantăm senzori în tot corpul pentru a da un substrat fizic emoțiilor.

[0] D.G. Dutton și A.P. Aaron, „Some Evidence for Heightened Sexual Attraction Under Conditions of High Anxiety”, Journal of Personality and Social Psychology 30 (1974): 510 – 517.

Probabil că primele dispozitive dotate cu conștiință artificială nu vor avea emoții convingătoare. Aceste mașinării ar putea imita tonul afectiv al vocii, dar captarea adevărului profund al emoțiilor, modalitatea în care sunt reprezentate în creierul uman se află departe în viitor. Mai întâi trebuie să se răspundă unor întrebări științifice de bază. Clișeul hollywoodian al androidului lipsit de emoții s-ar putea să se dovedească corect, cel puțin pentru o vreme.

A patra și ultima componentă pe care trebuie s-o integrăm în mașinăria noastră conștientă e un motor de căutare care să vorbească. Vrem o mașinărie care poate vorbi cu noi despre experiențele ei conștiente. În principiu, vorbirea nu e necesară pentru conștiință, dar cred că scopul conștiinței artificiale, după părerea majorității oamenilor, e o mașinărie cu o capacitate asemănătoare oamenilor de a vorbi și de a înțelege. Vrem să putem avea o conversație plăcută cu ea.

Ar putea părea că acest aspect al problemei a fost deja rezolvat, dat fiind că avem asistenți digitali ca Siri și Alexa. Aceștia pot înțelege întrebări verbale, pot căuta într-o bază de date și pot apoi să răspundă la întrebare. Nu e suficientă această capacitate dacă vrem o mașinărie care să vorbească despre stările ei interne? Dar problema e înșelătoare și spinoasă. Siri operează în principal în domeniul limbajului. Îi dai cuvinte, caută mai multe cuvinte pe Internet și apoi îți răspunde în cuvinte. Dacă întrebi care e cel mai apropiat restaurant, Siri nu știe ce e un restaurant, îl percepe ca parte a unui amestec statistic de cuvinte. Prin opoziție, creierul uman poate traduce vorbirea în informație nonverbală și apoi din nou în vorbire. Dacă cineva te întreabă: „Ce gust are lămâia în comparație cu o portocală?”, nu răspunzi ca un motor de căutare pe Internet. Nu te bazezi pe asociații de cuvinte. Traduci vorbirea în informații gustative și compari cele două gusturi pe care ți le amintești, apoi traduci rezultatul în cuvinte pentru a oferi răspunsul. Acest dialog rapid între vorbire și

multe alte domenii informaționale este incredibil de dificil de realizat artificial. Din câte știu, problema nu a fost încă rezolvată într-o modalitate sistematizată sau generală. Google poate, într-o anumită măsură, să traducă informația din imagini vizuale în cuvinte, dar mașinăria noastră conștientă ar trebui să coreleze informații din toate domeniile imaginabile.

Având în vedere toate promisiunile și toate dificultățile despre care tocmai am vorbit, cât de aproape suntem să construim mașinării conștiente?

Primele dispozitive înzestrate cu conștiință vizuală – mașinării care au și atenție, și o schemă a atenției care să proceseze informație vizuală – ar putea fi construite în următorii zece ani. Cred că va dura mai mult până când mașinăriile vor avea o gamă foarte extinsă de conținuturi, similară celei a conștiinței umane. E un proiect pe termen lung construirea unei mașinării capabile să vadă, să audă, să guste, să atingă, să aibă gânduri abstracte și să construiască emoții, capabilă de o unică focalizare integrată a atenției, care să coordoneze domeniile din interior și din exterior, *și* care să poată vorbi articulat despre toate aceste conținuturi. Aș fi surprins dacă am ajunge la așa ceva în următorii 30 de ani (estimez realist că va dura în jur de 50 de ani). Mi s-a mai întâmplat însă să fiu surprins, iar tehnologia informației avansează într-un ritm uluitor.

Oamenii visează la automate inteligente încă de când există tehnologia mecanică. *Iliada* lui Homer, veche de aproape 3.000 de ani, menționează trepiede care se deplasau autonom, construite de zeul Hefaistos. În urmă cu 500 de ani, Leonardo da Vinci a conceput și a construit roboți mecanici pentru a-și distra protectorii[0]. în zilele noastre, toată lumea e familiarizată cu HAL 9.000 din filmul lui Stanley Kubrick *Odiseea spațială 2001* și cu C_3PO din *Războiul stelelor* al lui George Lucas. Un film mai recent, *Ex Machina* al lui Alex Garland, se concentrează exclusiv

[0] M.E. Moran, „The da Vinci Robot”, Journal of Endourology 20 (2006): 986 – 990.

asupra modului în care oamenii pot interacționa cu roboți inteligenți. Mintea umană e preocupată de mașinăriile conștiente de așa de multă vreme, încât ne e teamă, parcă, de cât de aproape suntem să le avem în viața reală. E una din situațiile în care e potrivită zicala: „Ai grijă ce-ți dorești”.

Încă nu avem un model cultural adecvat pentru a preconiza schimbările care vor surveni grație mașinăriilor conștiente. Operele științifico-fantastice fac adesea predicții corecte în legătură cu dispozitivele din viitor, dar efectul social al acestor dispozitive poate fi mai greu de imaginat. De exemplu, nimeni nu a prezis revoluția socială provocată de telefoanele mobile. La început am crezut că vor fi niște gadgeturi formidabile, cum sunt cuptoarele cu microunde, dar au ajuns să fie a treia emisferă a creierului uman. Au restructurat lumea noastră politică, economică și socială. Așa că singura predicție pe care pot s-o fac este că nimeni nu poate să prezică exact efectul social al mașinăriilor conștiente.

Lucrările științifico-fantastice încearcă să prezinte imaginea unui viitor în care există conștiință artificială, dar mă îndoiesc că realitatea va arăta la fel. În *Războiul stelelor*, mașinăriile conștiente nu par să pună vreo problemă etică serioasă. Sunt pur și simplu cetățeni de mâna a doua. Sunt niște servitori supuși și abili tehnic, sunt fermecători, dar dispensabili, și în general sunt tratați cu mai puțin respect decât animalele de companie. E oare o idee bună ca mașinăriile conștiente să fie tratate atât de neglijent, sau ar trebui să le acordăm drepturi morale? Celebra nuvelă *Omul bicentenar* a lui Isaac Asimov explorează această temă[0]. Faptul că putem construi mașinării conștiente ne dă dreptul să le controlăm și să le ucidem? Filmul *Blade Runner* (*Vânătorul de recompense*), bazat pe o nuvelă a lui Philip K. Dick, pune această

[0] I. Asimov, The Bicentennial Man (New York: Ballantine Books, 1976). [în ed. Rom., Povestiri ciberrobotice, antologie de Alexandru Mironov și Mihai Bădescu, Editura Științifică și Enciclopedică, București, 1986.]

întrebare etică[0]. Scriitori, regizori, oameni de știință și filosofi s-au străduit în zadar să răspundă la aceste întrebări.[0] S-ar putea să trebuiască să așteptăm până construim o conștiință artificială cu adevărat asemănătoare oamenilor înainte să putem dezvolta o etică pragmatică. Chestiunile vor arăta probabil altfel decât în lucrările științifico-fantastice, care sunt în esență o modalitate indirectă de a aborda dilemele din prezent. S-ar putea ca dilemele de mâine să nu semene deloc cu cele de azi.

Cel mai mare și mai apăsător mister nu e însă cum sau cât de rău vom trata noi roboții conștienți, ci cum ne vor trata ei pe noi. Roboții conștienți au devenit un clișeu apocaliptic. În *Terminator*, Skynet încearcă să ne omoare pe toți. În *The Matrix* suntem sclavii mașinăriilor. Dar, din nou, e vorba de lucrări științifico-fantastice, nu de realitate.

Poate fiindcă am fost supraexpuși la imaginarul științifico-fantastic avem tendința să înțelegem greșit conceptul de mașinărie conștientă. Majoritatea oamenilor au imaginea unui buton care este apăsat pentru a activa conștiința într-o mașinărie. Mașinăria s-ar gândi dintr-odată la ea însăși ca la un agent, o ființă independentă de restul lumii, care ar putea să-și urmărească propriile planuri. S-ar putea să înceapă să omoare oameni ca să înlăture competiția pentru resurse. Mașinăriile cu astfel de capacități ar fi o catastrofă pentru specia umană. Am însă o veste proastă. Mașinării superinteligente, autonome, care pot lua decizii și pot urma un plan există deja într-o formă incipientă și devin din ce în ce mai inteligente, din ce în ce mai repede. Nu au nevoie de conștiință ca să decidă că oamenii sunt niște obstacole care trebuie ucise. S-ar putea ca un camion autonom cu un plan prestabilit să te calce și să te omoare, fie că are sau nu o experiență subiectivă a faptului că te-a zdrobit sub roți.

---

[0] P.K. Dick, De Androids Dream of Electric Sheep? (New York: Doubleday, 1968). [V. Ed. Rom., Visează androizii oi electrice? trad. De Ștefan Ghidoveanu, Nemira, București, 2017.]
[0] D. Levy, „The Ethical Treatment of Artificially Conscious Robots", International Journal of Social Roboticst (1929): 209 – 216.

Pentru mine, ca om de știință care cercetează conștiința, întrebarea este: ce se întâmplă dacă mai adăugăm o piesă – un model al sinelui care îi spune mașinăriei că are experiență subiectivă și că și alții pot avea conștiință la rândul lor? Cum schimbă asta ecuația socială?[0] Cred că această perspectivă ne dă speranță în privința viitorului nostru alături de mașinării inteligente.

Oamenii sunt animale întru totul sociale. Este ceea ce ne definește ca specie. Dar nimeni nu poate fi competent din punct de vedere social fără o înțelegere profundă a conștiinței și fără o capacitate intuitivă de a o atribui celorlalți oameni. Această capacitate ne permite să recunoaștem alte minți, să ne înțelegem reciproc gândurile și senzațiile și să dăm răspunsuri bine calibrate. E liantul care ține oamenii laolaltă. E rădăcina cooperării și a societății.

Imaginează-ți o lume în care oamenii și-au pierdut această cunoaștere profundă, bazată pe modele. În schimb, am fi programați altfel. Nu am fi inerți și fără reacții, ca oamenii cu leziuni cerebrale. Am putea funcționa în continuare ca agenți autonomi, am putea să învățăm, să acționăm conform unui plan și să luăm decizii inteligente, dar toate acestea în lipsa unui construct al conștiinței. Am deveni entități independente, fiecare cu propriile planuri, tratându-i pe ceilalți ca pe niște obiecte. N-am putea să cooperăm în vreun fel semnificativ, chiar dacă am avea inteligența de a ști că această cooperare ar fi avantajoasă; nu am înțelege că și ceilalți au minte, așadar nu am putea să identificăm stările mintale ale celorlalți și nu ne-am putea coordona gândurile, scopurile și acțiunile. Să ucidem ar deveni la fel de banal ca îndepărtarea unui obstacol, dat fiind că nu am mai înțelege valoarea celorlalte minți. Am

[0] B. Hood, The Self Illusion: How the Social Brain Creates Identity (Oxford, UK: Oxford University Press, 2012); F. Podschwadek, „De Androids Dream of Normative Endorsement? On the Fallibility of Artificial Moral Agents”, Artificial Intelligence and Law 25 (2017): 325 – 339; J. Sullins, „Artificial Phronesis and the Social Robot”, Frontiers în Artificial Intelligence and Applications 290 (2016): 37 – 39.

trăi într-o lume de monștri inteligenți, centrați pe scop.

Aceasta este lumea mașinăriilor inteligente lipsite de conștiință, și o construim chiar acum. Aș prefera să trăiesc într-o lume de mașinării care știu ce e conștiința și o pot atribui oamenilor, tot așa cum sunt recunoscător că trăiesc într-o lume în care oamenii își pot atribui conștiință unii altora.

Nu susțin că adăugarea conștiinței ar fi vreun ingredient magic care ar face ca mașinăria să se comporte etic. Știm cu toții că în lume sunt oameni plini de ură, printre care agresori și sociopați. Toți au conștiință și probabil știu că și oamenii cărora le fac rău sunt înzestrați cu conștiință la rândul lor. Dar ei sunt niște piese puse anapoda într-un mozaic mai amplu. Sunt întotdeauna minoritari, oamenii fiind în mod fundamental o specie prosocială. De fiecare dată când intru într-un supermaket, mă bazez pe faptul că oamenii din jurul meu sunt mai mult sau mai puțin încadrați în matricea socială. Evoluția ne-a oferit o soluție imperfectă, dar destul de bună din punct de vedere statistic, la problema coeziunii sociale. Ceea ce ține omenirea laolaltă, ceea ce ne deosebește de specii mai puțin cooperante și mai antagonice, este gândirea noastră hipersocială. Această capacitate are probabil multe surse, dar nu ar fi posibilă fără un model intern care să ne spună că avem conștiință și că și ceilalți au conștiință.

Dat fiind că lumea se grăbește să creeze mașinării cu inteligență și autonomie asemănătoare celor umane sau chiar supraumane, cred că ar fi indicat să le atribuim și niște trăsături umanizante, astfel încât să aibă o șansă de a se integra printre noi. De ce să nu aplicăm aceeași soluție a cooperării pe care evoluția a găsit-o pentru noi? Dacă teoria schemei atenției este corectă, atunci construirea conștiinței artificiale este unul dintre cei mai folositori pași pe care îi putem face pentru a reduce riscurile tehnologice care ne pândesc în viitor.

9
Transferul minților[0]

Nu vreau să trăiesc veșnic. Întotdeauna am considerat că, dacă viața are un scop, acesta trebuie să fie legat de contribuția fiecăruia la societate, iar viața veșnică pare să încline balanța mai degrabă spre a lua decât spre a dărui. Și ar fi și plictisitoare. Ca într-un joc video în care nu mori niciodată, nu există niciun pericol, iar sufletul e cuprins de un fel de apatie. Așa văd eu lucrurile acum. S-ar putea să-mi schimb perspectiva pe măsură ce mă apropii de sfârșit. Adevărul este că mulți oameni, poate cei mai mulți, vor să trăiască pe termen nedefinit, iar tehnologia se îndreaptă spre acest deziderat. Nu mă refer la nemurirea asistată medical, care nu pare plauzibilă fizic. Mă refer la ceva ce se numește transferul minții (*mind uploading*) – trecerea esenței unei persoane, a conținuturilor sale cerebrale, pe o platformă artificială. Mai devreme sau mai târziu, oamenii vor inventa o viață de apoi în care mintea umană va trăi pe termen nedefinit într-un univers simulat și va interacționa cu lumea reală prin videoconferințe sau accesorii robotice. Viața e ciudată, dar, dintre toate lucrurile ciudate pe care le-am întâlnit sau la care m-am gândit, această viață de apoi computerizată e probabil cea mai bizară.

Dacă teoria la care am lucrat în ultimii zece ani e corectă, atunci tot ce ține de minte – amintirile, emoțiile, personalitatea, chiar conștiința mea – e un produs al mecanismelor cerebrale fizice și se poate copia. Ar trebui să putem scana creierul suficient de amănunțit pentru a putea crea o simulare a acestuia, un dublu artificial al informațiilor și algoritmilor săi, care ar putea continua să trăiască după moartea biologică a persoanei. Prin această afirmație nu încerc să minimalizez complexitatea umană. Așa cum am mai spus, mintea este o sculptură de bilioane de fibre de informație, în permanentă schimbare și minunat

[0] Titlul în engleză e *Uploading Minds*. *To upload*, în acest context, înseamnă încărcarea unui program sau a unor date pe un suport digital (n. Red.).

de complicată. Dar nimic din ea nu e atât de misterios încât, în principiu, să nu poată fi copiat într-un alt dispozitiv de procesare a informației, așa cum ai copia un fișier dintr-un computer în altul.

Sunt mulți cei care cred că tehnologia transferului minții (*mind, uploading*) e foarte aproape.1 Eu nu sunt la fel de optimist în privința cronologiei. Deși e adevărat că tot ce ține de tehnologia computerelor avansează într-un ritm uluitor, ceea ce ține de înțelegerea sau scanarea biologiei cerebrale avansează cu pași mult mai mici. Cu toate astea, cred că e inevitabilă inventarea acestei tehnologii. S-ar putea să dureze secole sau s-ar putea să apară mai devreme decât ne-am aștepta, dar tendințele tehnologice și motivația umană se îndreaptă în direcția asta.

Această provocare este fundamental diferită de cea care presupune construirea conștiinței artificiale. Nu e necesară aceeași cunoaștere a modului în care funcționează integrat diferitele componente ale creierului, dat fiind că nu trebuie să construiești conștiința de la zero. Nu trebuie decât să copiezi un creier care există deja. Nu e nevoie să știi de ce conexiunile sunt așa cum sunt, atâta timp cât poți realiza o copie perfectă.

Provocarea constă în faptul că pentru copierea unui creier e nevoie de o scanare aproape inimaginabil de amănunțită.

Ca să construim o mașinărie care să poată transfera minți, trebuie mai întâi să descoperim seria minimă de date cerebrale care conțin esența persoanei. Majoritatea specialiștilor în neuroștiințe consideră că operațiile cerebrale sunt realizate în principal de neuroni interconectați prin sinapse, contactele specializate care îi permit informației să circule într-un mod controlat de la un neuron la altul. Creierul uman conține 86 de miliarde de neuroni.2 E posibil să conțină 100 de trilioane de sinapse, poate chiar de zece ori mai multe.

Așa-numita teorie a neuronilor, principiul conform căruia funcționarea creierului se realizează prin neuroni și sinapse,

a fost descoperită cu ceva mai mult de un secol în urmă de omul de știință spaniol Santiago Ramón y Cajal.3 A fost unul dintre geniile neuroștiințelor, iar pentru cercetările lui a primit în 1906 Premiul Nobel pentru medicină. A examinat la microscop bucăți de țesut cerebral folosind substanțe de contrast și a fost primul care a conturat o imagine corectă a modului în care funcționează creierul.4 Informația trece de la un neuron la altul prin delicatele dendrite și butoni terminali. Fluxul informației este îngrădit de sinapsele dintre neuroni, blocat aici sau deschis acolo, direcționat pe căi și circuite cerebrale specifice de la imput, prin procesare internă, la output. Această perspectivă de ansamblu este în esență aceeași cu a specialiștilor în neuroștiințe din prezent. Celulele cerebrale desenate cu măiestrie de Cajal apar și astăzi în manuale.

Inspirați de teoria neuronilor, oamenii de știință și inginerii au creat neuroni artificiali și i-au interconectat în rețele masive pentru a vedea cât de maleabilă sau de inteligentă poate deveni o imitație de sistem neural.5 Această tehnologie ne-a revoluționat lumea. Rețelele neurale artificiale s-au dovedit extraordinar de puternice și de adaptabile. Motoarele de căutare pe Internet, asistenții digitali care par să înțeleagă vorbirea, mașinile care se conduc singure, algoritmii de tranzacții de pe Wall Street, circuitele telefoanelor noastre inteligente – toate aceste lucruri, care au ajuns să fie ceva obișnuit în lumea noastră, funcționează parțial în virtutea rețelelor neurale artificiale.

Santiago Ramón y Cajal nu e doar părintele neuroștiințelor, ci și, într-un mod pe care nimeni nu l-ar fi putut anticipa, fondatorul revoluției tehnologice din prezent.

Principiul de la baza rețelelor neurale este că fiecare neuron este extrem de simplu, dar, când un număr mare de neuroni sunt conectați între ei, devin capabili de o mare putere de calcul. La nivel fundamental, un neuron nu face nimic mai mult decât să transmită un semnal. Să presupunem că neuronul A este conectat printr-o sinapsă la neuronul B. Când A se activează, transmite un semnal

electric care îl traversează, ajunge la sinapsă, sare prin ea folosindu-se de un mesager chimic și îl influențează pe B. Dacă sinapsa este excitatorie, semnalul amplifică activitatea electrică deja prezentă în neuronul B, crescând șansele ca acesta să-și transmită propriul semnal. Dacă sinapsa este inhibitorie, semnalul va diminua activitatea neuronului B, scăzând șansele ca acesta să-și transmită propriul semnal. Sinapsele pot să varieze în privința intensității, unele fiind capabile să transporte semnale mai puternice, permițându-i neuronului A să aibă un efect mai puternic asupra neuronului B, iar unele sunt mai slabe, neuronul A neavând aproape niciun efect asupra neuronului B. La un nivel foarte simplist, această influență sinaptică a neuronului A asupra neuronului B e tot ce se întâmplă, acțiune repetată de bilioane de ori într-o rețea interactivă masivă. Fiecare neuron primește imput de la un număr de până la 100.000 de alți neuroni. Fiecare neuron trebuie să numere semnalele pe care le primește și să ia o decizie: dată fiind salva de imput, toate semnalele excitatorii și inhibitorii, da-urile și nu-urile din acest moment, câștigă da-ul sau nu-ul? Dacă câștigă da-ul, neuronul își transmite propriul semnal pentru a influența rețeaua. Fiecare neuron trebuie să ia această decizie din nou și din nou. Din acest aparent haos de o simplitate repetitivă derivă operațiile complexe.

Rețelele neurale, biologice sau artificiale, se pricep excepțional de bine să învețe sarcini complexe. De exemplu, dacă vrei să înveți o rețea artificială să recunoască fețe, îi oferi imputuri pixelate de fețe de la o cameră digitală și îți dă înapoi informații despre identitatea feței. Între cele două acțiuni se află un labirint de neuroni și sinapse prin care trece informația. La început, rețeaua face o treabă proastă, asociind poate aleatoriu fețele cu numele. Dar de fiecare dată când încearcă primește un semnal de învățare. Pe măsură ce rețeaua învață, schimbă conexiunile dintre neuroni, stabilește cât de puternică e fiecare conexiune și dacă e excitatorie sau inhibitorie. La final, rețeaua învață sarcina ajustându-și tiparele sinaptice. Dacă

vede o imagine cu fața lui Jim, indiferent dacă e în umbră sau bine luminată, dacă zâmbește sau se încruntă, mașinăria asociază imputul vizual cu outputul corect și îți poate spune: „Acesta e Jim”. Nimeni nu știe dinainte care ar trebui să fie tiparele sinaptice corecte. Nu poți să programezi fiecare detaliu al unui dispozitiv bun de recunoaștere facială. În schimb, sistemul învață prin încercare și eșec până când ajunge la un tipar adecvat rezolvării sarcinii.

Grație ultimului secol de cercetare a neuronilor biologici și reușitelor recente din cercetarea rețelelor neurale artificiale, majoritatea specialiștilor în neuroștiințe consideră acum că esența creierului stă în tiparul de conectivitate a neuronilor. Conform acestei perspective, dacă am putea măsura toți neuronii din creierul cuiva, dacă am afla care neuroni sunt conectați între ei și dacă am cataloga aceste sinapse, am avea esența persoanei respective. Această hartă ipotetică a tuturor neuronilor și a conexiunilor sinaptice dintre aceștia se numește *conectom*,[6] cuvânt conceput prin analogie cu genomul. Ideea e că, dacă oamenii de știință au reușit să cartografieze genomul uman, o realizare considerată cândva imposibilă, atunci ar putea să rezolve o provocare tehnică și mai dificilă și să cartografieze conectomul uman. Fiecare om are propriul său conectom, care definește o minte unică.

În ultimii zece ani s-au realizat unele progrese în cartografierea conectomului. A fost publicat întregul conectom al unei specii de vierme nematod, *Caenorhabditis elegans*, și, mai recent, al musculiței de oțet.[7] Oamenii de știință pot de asemenea să decupeze o porțiune de câțiva milimetri din cortexul unui șoarece, să o înghețe, să o taie în secțiuni extrem de subțiri, să scaneze fiecare secțiune și să reconstruiască o mare parte dintre neuroni, straturile lor încrucișate și conexiunile lor sinaptice.[8] Această metodă nu a dus încă la reconstruirea tuturor sinapselor, dar conectomul complet al câtorva milimetri din creierul unui șoarece e, dacă nu chiar la îndemână, plauzibil în viitor.

Institutul Național de Sănătate sponsorizează în prezent

Proiectul Conectomului Uman, un efort global masiv al cărui scop ultim este cartografierea conectomului uman. În studiul creierului uman, metode de scanare RMN pot fi utilizate pentru a identifica rețele de conexiuni la un nivel din ce în ce mai amănunțit.9 Acest tip de scanare neinvazivă e puțin mai convenabilă pentru participant, dat fiind că nu presupune ca o parte a creierului subiectului să fie înghețată și feliată. Un voluntar poate sta într-un scaner RMN câteva ore și, gata! avem o scanare uluitor de detaliată. (Și eu mi-am scanat creierul de multe ori. E puțin inconfortabil, destul de plictisitor și am tendința să adorm, dar rezultatul e spectaculos.) Dar aceste așa-numite hărți ale conectomului uman sunt la o rezoluție mult mai slabă decât neuronii și sinapsele. Ilustrează tipare la scară mare, arătând cum o părticică din cortex de mărimea unei boabe de mazăre se conectează la o alta. Pe măsură ce avansează tehnologia de scanare, specialiștii în neuroștiințe se așteaptă ca interconexiunile cerebrale să poată fi măsurate la o rezoluție mult mai mare.

Progresele pe care le-am descris până acum în privința rețelelor neurale artificiale și a cartografierii conectomului creierului par la o primă vedere să dea motive de speranță în legătură cu transferul minții. S-ar zice că știm ce anume să măsurăm în creier și cum să simulăm neuronii. Dat fiind ritmul progresului, cu siguranță că vom putea realiza transferul minții în următorii zece sau douăzeci de ani, nu? Eu nu împărtășesc acest optimism. Vom ajunge acolo, dar nu curând, în următoarea secțiune voi explica de ce cred că suntem încă foarte, foarte departe de transferul minților.

Să ne mai uităm o dată peste cifrele pe care le-am menționat mai sus. Creierul uman conține aproximativ 86 de miliarde de neuroni și în jur de 100 de bilioane de sinapse, iar astea sunt numerele cele mai modeste.10 Nu știu să existe vreo tehnologie care să poată scana și măsura un asemenea volum. Dispozitivele actuale nu se ridică la acest nivel.

Scanerele RMN pot în prezent să măsoare creierul cu o rezoluție de aproximativ o jumătate de milimetru, ceea ce

este într-adevăr o realizare tehnologică impresionantă. Dar neuronii sunt mult mai mici, și sinapsele încă și mai mici. Pentru a detecta o sinapsă îți trebuie o scanare cu o rezoluție de un micrometru – o miime de milimetru. La această scară, s-ar putea să se vadă niște bulbi pe neuroni, despre care am putea presupune că sunt sinapsele.

Dar s-ar putea ca nici măcar această rezoluție să nu fie suficient de bună. Ar trebui să determinăm nu doar dacă bulbul acela care abia se vede e o sinapsă, ci și ce fel de sinapsă e și cât de puternic e semnalul pe care-l poate transporta de la un neuron la altul. O sinapsă mai mare tinde să lanseze un semnal mai puternic, deci, ca o primă metodă brută de măsurare a puterii sinaptice, ar trebui măsurată dimensiunea sinapsei. Ar fi necesară pentru aceasta o imagine clară, detaliată a fiecărei sinapse la o rezoluție mult mai bună decât cea de un micrometru. Dispozitivul de scanare ar trebui să identifice și dacă sinapsa este excitatorie sau inhibitorie. Pentru a descoperi această informație ar fi nevoie probabil de scanarea prezenței unui tip anume de moleculă în fiecare sinapsă. Nu există niciun fel de tehnologie pentru scanarea la acest nivel. S-ar putea ca prin morfologie – în funcție de forma specifică a sinapsei – să ghicim mai mult sau mai puțin precis despre ce fel de sinapsă e vorba, în loc să trebuiască să ne bazăm pe analiza chimică, dar chiar și pentru această scurtătură posibilă ar fi nevoie de o rezoluție incredibil de bună, poate de o miime de micrometru. Nu vorbim de îmbunătățirea tehnologiei RMN. Vorbim de un microscop electronic, în cazul studierii țesutului mort, sau, în cazul unui creier viu, de o nouă tehnologie de scanare, care nu a fost încă inventată.

Creierul conține sute, poate mii de tipuri diferite de sinapse.11 O joncțiune, de exemplu, implică o conexiune electrică directă între neuroni. E extrem de rapidă și de sigură și e esențială pentru funcționarea normală a anumitor părți ale creierului în care e necesară sincronizarea perfectă. Un alt tip de sinapsă funcționează ca un pulverizator, transmițând semnale care influențează

o rețea locală a creierului, nu doar un neuron din apropiere. Unele sinapse au mai multe tipuri de transmițători chimici, secretând câte unul în funcție de circumstanțe. Altele se pricep mai bine la schimburi rapide, parte a învățării pe termen scurt, în timp ce altele sunt mai stabile. Toate aceste tipuri diferite de conexiuni între neuroni, toate aceste nuanțe, inclusiv multe care probabil nu au fost descoperite încă, dinamicile lor diferite, viteza, și puterea, și adaptabilitatea lor ar trebui identificate pentru a putea fi construit conectomul.

Chiar dacă am reuși să scanăm toate aceste informații neuronale și sinaptice, tot ar rămâne de rezolvat problema celulelor gliale, adesea trecute cu vederea.12 Majoritatea specialiștilor în neuroștiințe se concentrează pe studierea neuronilor – de unde și numele profesiei. Dar creierul e plin de un alt tip de celule care depășesc numărul neuronilor într-un raport de 10 la 1. **Nevrogliile** au fost de mult trecute pe planul al doilea: sunt văzute strict ca niște celule de susținere, ca un schelet care dă formă creierului sau ca niște celule-servitori care îndeplinesc nevoile neuronilor și curăță reziduurile. Dar se dovedește că nevrogliile au proprietăți legate direct de procesarea informației. Parte din ele secretă substanțe chimice care influențează neuronii și sinapsele. Unele celule gliale chiar transmit aceleași tipuri de semnale electrochimice pe care le folosesc neuronii pentru a comunica. Funcțiile nevrogliilor nu sunt încă bine cunoscute, dar se știe că aceste celule nu sunt atât de diferite de neuroni din punct de vedere categorial, așa cum se considera mai demult. Tot ce se poate spune acum e că există mari lacune în ceea ce știm despre procesarea cerebrală, ceea ce ar trebui să ne îndemne la modestie.

Pe scurt, nu-ți recomand să fii printre primii care adoptă această tehnologie când se va inventa, pentru că o să aibă mare nevoie de îmbunătățiri.

Să presupunem că în viitor se inventează o tehnologie de scanare care poate identifica sinapsele din creierul tău. Versiunea beta poate lua în calcul doar neuronii, nu și

celulele gliale. S-ar putea să trebuiască să simplifice, să clasifice sinapsele în doar 100 de tipuri majore. S-ar putea să trebuiască să convertească puterea sinapselor – de exemplu, să permită 100 de amplificări posibile ale intensității, în loc să ilustreze o gradație mai nuanțată. Scanerul s-ar putea să ignore acțiunea hormonilor, care se răspândește difuz în întregul creier. S-ar putea să aibă o rată incredibilă de succes, înregistrând corect 99,99% dintre sinapse, și, cu toate astea, să rateze suficiente detalii încât să conteze. Rezultatul ar semăna întrucâtva cu originalul, dar nu putem ști ce minte defectă poate lua naștere din acest proces odată ce construiești un creier artificial pe baza acestor informații. Ar fi, poate, mahmură, bolnavă, disfuncțională afectiv și incapabilă să se concentreze?

Nu e greu să perturbi echilibrul normal al creierului. Chiar și o cantitate minimă din anumite droguri poate să provoace durere, confuzie, halucinații și convulsii. O comoție, care rupe fibre și provoacă inflamație, poate duce la luni sau chiar ani de gândire încețoșată și instabilitate emoțională. Chiar și micile iregularități pot face o mare diferență. Ar fi bine ca versiunea simulată a creierului tău să funcționeze al naibii de asemănător cu originalul, altfel experiența ei va fi un iad. Eu aș aștepta până la versiunea 1.000, când toate defecțiunile vor fi fost reparate. Porcii de Guineea n-o s-o ducă prea bine în povestea asta.

Dacă ne uităm la obstacole – volumul incredibil de informații despre creier care trebuie scanate, nivelul submicrometric de detaliu la care scanerul ar trebui să funcționeze, aspectele neuroștiințifice fundamentale care încă sunt necunoscute –, e tentant să renunțăm la întregul proiect și să presupunem că visul de a transfera minți e imposibil. Da, e absolut imposibil cu tehnologia de azi. Nici măcar versiunea îmbunătățită a tehnologiei de azi nu s-ar apropia la un an-lumină de transferul minților.

Dar în același timp sunt absolut sigur că se va realiza în cele din urmă. Oamenii se pricep să rezolve probleme tehnologice, în 1916, Einstein a prezis existența undelor

gravitaționale.13 Considera că efectul prezis va fi atât de absurd de mic, de vreo 10.000 de ori mai mic decât nucleul unui atom, încât tehnic va fi imposibil de confirmat. Nu-și putea imagina nicăieri în viitor o mașinărie cu un nivel suficient de sensibilitate.

Aproape 100 de ani mai târziu, mașinăria a fost construită și undele gravitaționale au fost confirmate.14 Presupunerea mea este că noua tehnologie va fi inventată, noi posibilități se vor deschide și transferul minții va deveni o realitate. Nu pot prezice când anume se va-ntâmpla acest lucru, pentru că depinde de inspirația oamenilor să inventeze mașinării necunoscute și în prezent de neimaginat. Dacă ar trebui să zic un număr, aș spune cel puțin 100 de ani, dacă nu mult mai mult, deși s-ar putea să mă înșel dacă cineva e lovit de inspirație mai devreme.

Și tot nu recomand să fii printre primii care o adoptă.

Transferul minții are două componente: prima constă în scanarea informației relevante din creierul pe care vrei să-l copiezi, după cum am explicat deja; a doua constă în crearea unei simulări funcționale a acelui creier. Să presupunem că problemele tehnice de scanare a creierului vor fi rezolvate. A fost inventat aparatul necesar pentru scanarea cerebrală și au fost asimilate toate detaliile creierului tău. Acum trebuie să folosim aceste date pentru construirea unei simulări funcționale a creierului.

Ar putea părea că a doua parte, simularea, este provocarea mai dificilă, dar, de fapt, este practic rezolvată. Piesele de hardware există deja. Neuronii artificiali și rețelele neurale sunt ceva standard. Adăugarea de noi tipuri de sinapse și de influențe subtile, cum ar fi efectele hormonale, nu este în sine dificil de simulat. Chiar și rețelele care conțin milioane de neuroni artificiali sunt ușor de construit. Companii din întreaga lume încearcă în prezent să construiască sisteme a căror complexitate rivalizează cu cea a creierului. Proiectul Creierului Albastru (*The Blue Brain Project*), de exemplu, folosește supercomputer pentru a simula ansambluri masive de neuroni asemenea celor din creier. Proiectul Creierului

Uman, Institutul Allen, Google Brain, Deep Mind, Cogitai și multe alte cercetări de pionierat se îndreaptă spre crearea de sisteme masive de rețele neurale artificiale. Construirea de rețele cu 86 de miliarde de neuroni și 100 de bilioane de sinapse nu e încă realizabilă cu tehnologia de azi. Dar tehnologia avansează rapid și, în special datorită progresului computerelor cuantice, nu am nicio îndoială că vom avea în curând mijloacele de a simula o rețea la scara creierului uman.

Acest progres masiv al tehnologiei este în parte ceea ce alimentează optimismul din prezent față de transferul minților. O piesă importantă din puzzle, piesa cea mai vizibilă, este deja rezolvată. Dar nu trebuie să uităm că o rețea neurală artificială la scara creierului uman, oricât ar fi de monumentală, nu e același lucru cu o minte transferată. Rețeaua în sine, fără tiparele potrivite de conexiuni între cei 86 de miliarde de neuroni, este o grămadă inutilă de numere. E ca și cum am construi o imprimantă care poate printa un creier artificial și materialul din care e compus, dar încă nu știm cum să măsurăm datele cerebrale pe care să le imprimăm, așa că nu putem printa decât un amestec inutil.

Ca să preiau o formulă a lui David Chalmers, pot să spun că avem o problemă cu adevărat grea în ceea ce privește transferul minții: cum să scanăm creierul suficient de amănunțit.

Să o luăm puțin înainte și să presupunem că am putea să-ți simulăm creierul. L-am scanat și am recreat rețelele neurale. Următorul pas e să introducem acest creier copiat într-un corp. Fără corp, nu e clar ce fel de experiență ar avea creierul tău artificial plutind într-o cuvă digitală.15 Dacă ar fi într-adevăr asemenea unui creier real, cred că ar fi dezorientat și și-ar pierde poate senzația sinelui, ale cărui baze sunt în corp. Știi unde ești fizic, unde sunt mâinile, și picioarele, și trunchiul tău, și asta îți oferă o ancoră primară.16 Fără această ancoră și fără contact cu lumea din jur, fără *încorporare*, presupun că ce-ai simți ar fi un fel de confuzie mintală asemănătoare celei induse de droguri.

Dar să hotărâm ce fel de corp artificial să-ți dăm: un robot fizic care poate să se plimbe prin lumea reală, sau un corp simulat care trăiește într-o lume virtuală? Un robot fizic e limitator. Mi se pare că implantarea unui creier simulat într-un corp muritor, fragil, ar rata cu totul potențialul de flexibilitate și longevitate al transferului minții.

Pe vremea când studiam în laboratorul meu modul în care creierul controlează mișcarea, am construit o mână umană simulată.17 Nu era o mână reală, robotică. Nu aveai cum să o atingi și nu putea să apuce obiecte din laborator. Era o mână virtuală sub formă de date computerizate. Tot ce vedeam era o matrice de numere pe un ecran. Având la bază scanările unei mâini reale, mâna virtuală avea tot ce-i trebuia. Fiecare os, fiecare tendon, fiecare mușchi. Avea forță musculară, vâscozitate, inerție și gravitație. Mușchii erau alcătuiți din fibre musculare distincte, unele cu contracție rapidă și altele cu contracție lentă. I-am atribuit mâinii neuroni senzoriali, neuroni motori alfa, beta și gama. Ar fi costat milioane de dolari și ar fi durat ani întregi dacă am fi încercat să construim un braț robotic atât de asemănător unei mâini reale, și s-ar fi putut să nu fi reușit niciodată. Însă mâna noastră virtuală nu a costat decât câteva mii de dolari pentru computer, iar tot procesul a durat câteva luni.

Dacă am putut simula o mână umană la mijlocul anilor 2000, cu resursele tehnologice limitate ale acelei perioade, atunci ar trebui să fie deja posibilă construirea unui corp virtual, realist, cu tot cu detaliile oaselor, mușchilor, nervilor și pielii. Nu știu dacă cineva a făcut asta – dar ar trebui să fie posibilă construirea unui avatar superrealist, ca dintr-un joc video, care seamănă întru totul cu un om.

După ce vom fi creat un creier simulat și un corp virtual, ultima noastră sarcină va fi să integrăm aceste elemente într-o lume virtuală. Putem să ne inspirăm din jocurile video în ceea ce privește lumile simulate, tridimensionale, cu propriile legi fizice. Tehnologia aproape a ajuns la un nivel corespunzător.

O lume virtuală cu adevărat convingătoare, cu detalii vizuale fine, cu sunete care se deplasează în spațiul virtual în moduri realiste, cu adierea vântului pe pielea virtuală, poate cu mirosuri și gusturi programate să interacționeze cu papile gustative și căi respiratorii – o lume virtuală atât de detaliată încă nu există. Dar ar putea exista. Nu trebuie inventată vreo tehnologie nouă. Limita e impusă de puterea de procesare. Dacă toate supercomputerele și-ar uni forțele, îmi imaginez că ar reuși să simuleze un creier uman, corpul său și un apartament realist de trei camere, în care ființa artificială să se poată relaxa. Integrarea unui număr mare de minți într-un mediu mai larg ar necesita o creștere masivă a puterii de procesare.

Îmi dau seama că imaginea pe care o zugrăvesc aici despre transferul minții este bizară și neuniformă. O parte din tehnologie există deja. Altă parte urmează să fie inventată, probabil în următorii zece ani. Computerele au nevoie de mai multă putere de procesare – și în curând o vor avea. Alte tehnologii sunt însă departe în viitor, probabil la sute de ani distanță. Elementele problemei care țin exclusiv de tehnologia informației sunt mult mai aproape de realizare, iar elementele legate de biologia cerebrală se vor dezvolta mai lent. Dar nu văd nimic care să se opună tehnologiei transferului minții. Mai devreme sau mai târziu, probabil mai târziu, mințile oamenilor vor fi preluate din creierul biologic și transferate într-un format digital.

Nu am o perspectivă distopică asupra transferului minții. Implică riscuri majore, dar cred că prezintă și mari posibilități. Noi, oamenii, în felul nostru dezordonat, avem tendința să ne dăm seama ce funcționează și ce nu, și cred că transferul minților va aduce un haos etic și cultural care în cele din urmă se va aranja de la sine. Voi menționa aici cinci riscuri potențiale, înainte de a spune care cred că sunt unele dintre beneficiile transferului minții.

Riscul potențial 1. În climatul social de astăzi, dat fiind că schimbăm dispozitivele tehnologice o dată la câțiva ani, transferul minții nu ar fi practic. Mintea ta transferată ar fi

norocoasă să reziste zece ani, până la apariția unui nou sistem de operare, cu care nu va mai fi compatibilă. Ar trebui s-o aruncăm pe bunica așa cum am aruncat WordStar, programul de procesare de text din anii 1980 care acum e atât de învechit, încât sunt doar câțiva cei care-și amintesc că a existat. Abordarea capitalistă, consumeristă, cu o rată scăzută de retenție pe care o avem în domeniul tehnologiei informației ar trebui să se schimbe semnificativ înainte ca o platformă de transfer al minții să poată într-adevăr să prelungească viața cuiva.

Riscul potențial 2. Creierul uman are o capacitate de memorie extrem de mare, care nu a fost niciodată atinsă, deși teoretic e limitată. Pe măsură ce un creier simulat ar stoca amintiri, sinapsele s-ar reconfigura și în cele din urmă ar rămâne fără capacitatea de a depozita noi amintiri fără să le fragmenteze pe cele anterioare. Nu cred că știe cineva care e capacitatea maximă sau când s-ar putea atinge limita. Se poate să dureze sute de ani. Poate în cele din urmă tehnicienii vor descoperi cum să adauge sinapse anumitor structuri cerebrale despre care se știe că sunt implicate în memorie, cum ar fi hipocampul, pentru a-i oferi din când în când minții simulate niște spațiu de memorie în plus. Dacă nu, mintea simulată va trebui să se descurce cu o fereastră de memorie în care evenimentele din ultimele câteva sute de ani rămân clare, pe când tot ce a fost până atunci se va estompa treptat.

Riscul potențial 3. Care ar fi drepturile acordate minților virtuale și celor biologice? 18 Pentru ca această tehnologie să funcționeze, versiunea simulată a cuiva va fi supusă unei torturi existențiale, iar și iar, până ce metoda va fi pusă la punct. Dacă mintea de probă e aproape de ceea ce se dorește, dar nu e chiar versiunea ultimă, e etic să o distrugi și să încerci din nou? Dacă faci mai multe copii ale aceleiași persoane, asta înseamnă că fiecare copie e mai puțin valoroasă și mai dispensabilă decât cealaltă, sau îi sunt acordate fiecăreia drepturi morale? Și cât de importantă mai e persoana inițială, biologică, atâta timp cât se păstrează o versiune a minții ei? Sau, cu alte cuvinte, ce se

întâmplă cu sanctitatea vieții și a individului când există deja trei copii ale aceleiași persoane?

Riscul potențial 4. În multe religii, viața de după moarte e văzută ca recompensă pentru respectarea regulilor. De această putere se abuzează adesea. Promisiunea vieții de apoi a dat un imbold esențial cruciadelor sângeroase din Evul Mediu. Atacatorii sinucigași din zilele noastre sunt ademeniți cu recompense edenice. Dar liderii care mizează pe glorioasa viață de dincolo sunt dezavantajați: produsul lor nu e palpabil. Imaginează-ți ce putere de coerciție ar avea viața de după moarte dacă ar fi verificabilă obiectiv. Ai putea să-ți iei telefonul mobil și să vorbești sau să schimbi mesaje cu oamenii care sunt deja acolo, ai putea să afli dacă le place și chiar să vezi ce notă i-au dat pe Yelp. Cei ce vor controla accesul la această viață palpabilă de după vor controla lumea. Aproape toți oamenii de pe Pământ ar trece peste orice obstacol, chiar și peste cele dubioase din punct de vedere etic – mulți oameni, chiar și peste cele evident imorale –, pentru o șansă la nemurire. Mi-e ușor să-mi imaginez cum noua tehnologie va ajunge foarte rapid să servească unor scopuri politice întunecate.

Riscul potențial 5. Ale cui minți vor fi transferate? Ale bogaților? Ale oamenilor străluciți? Ale politicienilor puternici?

Pe principiul „primul sosit, primul servit”? S-ar putea ca resursele să fie drastic limitate, ceea ce ar duce la o competiție foarte complicată din punct de vedere etic. Sau s-ar putea ca memoria și puterea de procesare să ajungă atât de ieftine în viitorul îndepărtat încât asta să nu mai fie o problemă. Poate că platforma va fi un fel de Youtube, unde oricine se poate înscrie. Poate că va fi cumva neutră, ca Internetul, sau poate că unii oameni vor avea o lărgime mai mare de bandă. Sper ca oamenii ale căror minți vor fi transferate să fie buni și să reprezinte o sursă de inspirație, pentru că, dată fiind durata extinsă a vieții lor, vor avea probabil o influență însemnată asupra restului societății.

Chiar presupunând că totul merge bine și că sunt

evitate riscurile potențiale, rămâne totuși greu de preconizat ce ar putea însemna transferul minții la nivel filosofic sau cultural. Să ne gândim la cea mai simplă întrebare: ai fi cu adevărat tu?

Îți dorești să nu mori, așa că într-o zi te duci la o clinică de transfer, stai cinci ore într-un scaner, în timp ce lumini și sunete se aprind și îți bâzâie pe la urechi, și ieși de-acolo puțin amorțit pe la încheieturi, dar altfel bine. Hai să fim generoși și să ne prefacem că tehnologia funcționează perfect. A fost testată și s-au eliminat toate erorile. Îi oferă minții tale un corp virtual standard, suficient de confortabil, căruia i se atașează fața și vocea ta, într-un mediu virtual adecvat. Să zicem că se întâmplă toate astea.

Cine este acest al doilea tu?

Primul tu, să-i spunem tu cel biologic, a plătit o avere pentru această procedură. Și cu toate astea pleci de la clinică la fel de muritor cum ai intrat. Ești tot o ființă biologică și în cele din urmă o să mori. Conduci spre casă și te gândești: „Asta da risipă de bani”.

În același timp, versiunea ta simulată se trezește într-un apartament virtual și se simte ca și cum ar fi tot tu. Are continuitatea experienței. Își amintește că a intrat în clinică, a plătit cu cardul, a semnat un formular și s-a întins pe masă. Se simte ca și cum ar fi fost anesteziat, după care s-a trezit în altă parte. Creatura asta are amintirile tale, personalitatea ta, tiparele tale cognitive și înclinațiile tale afective. Se dă jos din noul pat și spune: „Uau, a funcționat! Nu-mi vine să cred că sunt aici! Asta da investiție”.

Nu vreau să-i mai spun „creatura asta”, fiindcă această minte este o versiune a ta. Să-i spunem tu cel simulat. „Simularea” ta decide să exploreze. Ieși din apartament în lumina soarelui într-o zi perfectă și descoperi o versiune virtuală a orașului New York. Sunetele, mirosurile, priveliștile, oamenii, senzația trotuarului sub tălpi, toate sunt prezente, însă cu mai puține gunoaie, iar șobolanii sunt complet igienici și adăugați pentru culoare locală. Intri în vorbă cu diverși trecători, așa cum n-ai îndrăzni niciodată în adevăratul New York, de teamă să nu primești un pumn

în față. Te oprești la o cafenea și te așezi la o masă cochetă de fier forjat de pe trotuar ca să savurezi un latte. Nu are gustul la care te așteptai. Nu ai deloc senzația că ceva ți-ar ajunge în stomac. Și nici nu ajunge, pentru că nu e un lichid real și tu nu ai stomac. Îți dai seama că n-o să mai trebuiască niciodată să te duci la toaletă. Degetele tale nu au amprente – sunt netede, ca să nu se risipească memorie pe detalii minore. Să respiri nu e tocmai cum știai. Dacă ți-ai ține respirația, nu ai ameți, fiindcă nu există oxigen în această lume virtuală. Descoperi că ești echipat cu un telefon mobil inteligent și suni la numărul care era al tău – numărul pe careți amintești că-l aveai cu doar câteva ore în urmă, când ai intrat în clinică.

Tu cel biologic răspunde la telefon.

— Hei, spui tu cel simulat, sunt eu, adică tu. Care-i treaba?

— Sunt deprimat, nenorocitule. Asta-i treaba. Sunt acasă și mănânc înghețată. Nu-mi vine să cred că am cheltuit atâția bani pe nimic.

— Pe nimic? Hai, frate! Nu ți-ar veni să crezi cum e aici! Pare cam plictisitor la prima vedere, dar sunt sigur c-o să mă distrez. Am trecut pe lângă un cinema și o librărie, și aici avem bani, așa că pot să fac cumpărături, slavă cerului; se spune că simulatorul de *Războiul stelelor* e atât de real, că te simți exact ca-n film și poți chiar să fii Wookie-ul. Și îl mai știi pe Kevin, tipul care a murit de cancer săptămâna trecută? Și el e aici! E bine și are aceeași slujbă. Vorbește prin Skype cu vechiul lui studio de yoga de trei ori pe săptămână ca să țină cursul de fitness. Dar prietena lui din viața reală l-a părăsit pentru cineva care încă n-a murit, așa că e puțin trist. Are însă cu cine să se vadă aici dacă vrea.

O să mă opresc aici, deși umorul situației e irezistibil. Dincolo de detaliile amuzante, rămâne o dilemă filosofică foarte reală, cu care oamenii vor trebui să se confrunte în cele din urmă. Care va fi relația dintre tu cel biologic și tu cel simulat?

Prefer o abordare geometrică a situației. Să ne imaginăm că viața ta e ca partea de sus a literei Y. Te naști

la bază, și, pe măsură ce crești, mintea ta este modelată și modificată de-a lungul unei traiectorii. Apoi te duci să te scanezi și, din acel moment, Y-ul se bifurcă. Există acum două traiectorii, fiecare un tu legitim. Să spunem că partea din stânga ești tu-cel simulat și partea din dreapta ești tu-cel biologic. Partea din tine care trăiește nelimitat e reprezentată de trunchiul literei Y și de partea din stânga sus. Așa cum sinele tău din copilărie continuă să facă parte din sinele tău biologic când ești adult, tot așa trunchiul din litera Y continuă să trăiască în șinele simulat. Odată scanarea finalizată, cele două ramuri ale literei Y o apucă pe căi diferite, acumulând experiențe diferite. Partea din dreapta o să moară. Tot ce i se întâmplă *după* punctul de bifurcație nu va ajunge la nemurire – asta dacă nu alegi să te scanezi din nou, ceea ce ar duce la apariția unei a treia ramuri, geometria complicându-se și mai mult.

Ce ia naștere aici nu e un tu unic, ci o versiune complexă topologic, un hipertu cu două sau mai multe ramuri. Una dintre acestea o să fie muritoare, iar celelalte vor avea o durată de viață nedefinită, în funcție de cât timp va fi menținută platforma virtuală.

Poate crezi că, din moment ce tu-cel biologic trăiești în lumea reală și tu-cel simulat trăiești într-o lume virtuală, nu vă veți întâlni niciodată, așa că n-o să apară nicio complicație. Dar cine se mai întâlnește față în față în zilele noastre? Oricum interacționăm în principal prin media electronice. Tu-cel simulat și tu-cel biologic sunteți două ipostaze pe deplin funcționale, interactive, capabile ale persoanei tale, aflate în competiție în același univers social și economic interconectat. Și se prea poate să vă regăsiți într-o conferință video sau prin vreo altă tehnologie.

Presupun că o soluție la această confuzie existențială ar fi stocarea copiei tale cerebrale fără activarea simulării. O dată la ceva vreme poți să te întorci la clinică ca să-ți scanezi din nou creierul. Dacă mori într-un accident, sau de boală, sau de vârstă înaintată, atunci cea mai recentă copie – sau copia pe care o specifici în testament – poate fi activată și transformată într-o simulare vie.

O altă modalitate, mai degrabă distopică, de a controla numărul de copii virtuale ar fi o lege care să impună ca, odată ce mintea ta a fost transferată pentru prima dată și simularea ta a fost activată, corpul tău biologic să fie ucis și transformat în fertilizator. Am avea atunci moguli care ar da mită ca să încalce legea. Ei o să vrea cinci sau zece copii ale propriei persoane, plus originalul. Am avea nevoie de un agent special, ca într-un film de la Hollywood de mâna a doua, a cărui treabă ar fi să găsească și să omoare copiile ilegale. Lucrurile se complică rapid.

S-ar putea și ca sistemul să rămână haotic și nereglementat, fără nicio restricție a numărului de copii, ceea ce ar duce la revoluționarea concepției noastre despre identitate și individualitate.

La cel mai simplu nivel, transferul minții poate conserva oamenii într-o viață de apoi pe termen nedefinit. La masa de familie de Crăciun s-ar putea alătura bunica prin videoconferință, cu tableta proptită într-o parte a mesei – presupunând că mai are timp de familia ei biologică cu toate posibilitățile oferite de locul de joacă simulat. Poate că și-a ales un avatar atletic și e ocupată cu cățăratul pe munți virtuali.

La acest tip de viață de apoi idealizată se gândesc oamenii când își imaginează beneficiile transferului minții. E un rai făcut de oameni. Dar, spre deosebire de raiul tradițional, nu e o lume separată. E conectată direct la lumea reală. Să ne gândim la modul în care interacționăm acum cu lumea. Dacă ai un stil de viață tipic zilelor noastre, atunci interacțiunea cu oamenii din spațiul fizic din jurul tău e doar o mică parte din viața ta. Legătura ta cu lumea este aproape exclusiv digitală. Ai acces la știri printr-un ecran sau prin căști. Locuri îndepărtate sunt reale pentru tine în principal pentru că afli despre ele prin media electronice. Politicieni, vedete, chiar și unii prieteni și membri ai familiei au doar o prezență virtuală în viața ta. Oamenii lucrează frecvent în birouri unde își cunosc colegii doar prin transmisiuni video și text. Se prea poate ca fiecare dintre

noi să trăim deja într-o lume virtuală, cu un flux constant de informații venite prin CNN, Google, Youtube, Facebook, Twitter și mesaje-text. Trăim într-un multivers ciudat, fiecare dintre noi în bula sa virtuală, care se poate întrepătrunde uneori cu altele și care mai apoi se separă în spațiul real, toate fiind conectate însă prin rețeaua socială globală. Dacă viața virtuală de apoi va deveni realitate, oamenii de acolo, cu aceeași personalitate și aceleași nevoi pe care le aveau în viața biologică, nu vor avea niciun motiv să se izoleze de restul.

Lucrurile nu se vor schimba prea mult pentru ei. Lumea socială, politică, economică, virtuală și reală, s-ar conecta într-o civilizație mai mare, aflată în continuă expansiune. Lumea virtuală ar putea la fel de bine să fie un alt oraș de pe Pământ.

Timp de zeci de mii de ani, oamenii au transmis cunoașterea de la o generație la alta spunând povești despre trecut. Venerarea înaintașilor a reprezentat, fără îndoială, o parte activă a acestui proces.19 Menținând vie ideea că strămoșii noștri sunt în jurul nostru și ne veghează, ne susțin nevăzuți, oamenii au putut păstra o parte din înțelepciunea lor ca să-i călăuzească în viață.

Mă îndoiesc că atunci când scrisul a fost inventat în Sumer20, în urmă cu vreo 6.000 de ani, ar fi putut prezice cineva cât de mult ne va transforma specia. Inițial a fost un instrument economic, o modalitate de a nota cât s-a vândut din ce. Dar scrisul a realizat ceva esențial. Le-a permis oamenilor să vorbească direct peste generații. S-au extins dintr-odată volumul și precizia informației care putea fi acumulată în timp. Mai ales știința și tehnologia depind de acumularea de informație extrem de precisă, înregistrată de savanții de dinainte. Dar același lucru se poate spune mai mult sau mai puțin despre știința economică, teoria politică, ideologia religioasă, tendințele artistice și despre orice alt aspect al existenței noastre. Fără scriere, civilizația modernă nu ar fi fost posibilă, pentru că nu s-ar fi format la fel, prin adăugarea unui nou strat peste cele din trecut.

Transferul minții ar putea conserva mai multă informație, cu un nivel mai mare de fidelitate. Ar extinde și tipurile de informații care ar putea fi transferate – nu doar fapte simple care pot fi scrise într-o carte, ci și detalii de personalitate și abilități care transpar mai ușor atunci când vorbești direct cu cineva. Într-un fel ciudat, ne-am putea întoarce la ceva asemănător venerării strămoșilor. Dar, în loc să păstrăm o amintire vagă a înțelepciunii și faptelor de vitejie ale înaintașilor, prin tradiția orală, ei ar fi aici să ne vorbească direct. Și n-ar fi nevoie să ne șoptească sfaturi înțelepte de dincolo de abis. Ar putea să contribuie activ la societate prin media sociale. Din perspectiva mea, această modificare a felului în care informația se transmite în timp e probabil adevărata sursă a revoluției ascunsă în noua tehnologie. Transferul minții ar modifica specia noastră mai mult chiar decât au făcut-o evoluția limbajului și inventarea scrisului. Însă această schimbare s-ar putea să nu fie întotdeauna constructivă. La fel ca în cazul scrierii, televiziunii sau Internetului, când societatea adoptă o nouă tehnologie care crește fluxul de informație, această schimbare aduce inevitabil o creștere a dezinformării și a memelor sociale negative.

Pentru a reduce riscurile, am decide poate să conservăm înțelepciunea celor mai realizați oameni. Un Einstein e diferit de o persoană obișnuită pentru că are un set deosebit de legături sinaptice – o combinație fericită între ce a primit la naștere și ce a învățat mai târziu. Se întâmplă ceva asemănător când construiești și instruiești rețele artificiale. Începi cu un număr mare de rețele similare făcute din același număr de neuroni, le instruiești și vezi ce se întâmplă. Unele vor deveni genii care îndeplinesc extrem de bine sarcina dorită, iar altele se blochează, în mod dezamăgitor, în stări suboptime. Rețelele geniale nu au nevoie de informații diferite care să le definească. La urma urmei, sunt făcute din același număr de neuroni. Au ajuns cumva la un tipar bun de conexiuni sinaptice. Și, cu toate astea, nu putem identifica sinapsele bune. Nu putem spune: „Aha, această sinapsă inhibitorie a

ajuns la puterea optimă de conexiune. Ce frumusețe, ce geniu!” Nu, tehnicianul nu știe de ce anume rețeaua funcționează atât de bine. Prin instruire și prin întâmplare, sistemul a ajuns la puterea sinaptică optimă. Asta e rețeaua pe care vrei să o salvezi. La nivelul cel mai pragmatic, transferul minții ar conserva rețelele potrivite. Ar salva oameni care au găsit o serie bună de conexiuni ce susțin o serie utilă de abilități. Toată lumea capătă diverse abilități de-a lungul vieții. Imaginează-ți cum ar fi dacă am putea să conservăm aceste abilități.

Muzica ar fi evoluat diferit dacă Mozart ar mai fi trăit 200 de ani. Sau Beethoven. Sau, de ce nu, Elvis sau John Lennon. Nu spun că schimbarea ar fi fost bună sau rea, dar presupun că muzica s-ar fi schimbat mai puțin în timp. Istoria ne spune că adevărata revoluție în muzică tinde să aibă loc odată cu schimbarea generațiilor, când vechea gardă trece în plan secund și cea nouă încearcă să schimbe starea de fapt. Dacă ar fi existat transferul minții, pesemne că vocile noi nu ar fi apărut niciodată, iar vechile stiluri ar fi atins înălțimi nebănuite.

Să ne imaginăm cum ar fi evoluat limba engleză dacă Chaucer sau Shakespeare ar fi continuat să ne vorbească, să ne învețe și să scrie. Majoritatea vorbitorilor moderni de engleză îl găsesc pe Shakespeare dificil și pe Chaucer neinteligibil. Limbile se schimbă în timp într-un ritm studiat îndeaproape de lingviști. Se presupune că acum nu mai mult de 6.000 de ani un grup de oameni vorbea o limbă dispărută, numită protoindo-europeană, din care s-au desprins numeroase limbi moderne.[0]21 Dar schimbările lingvistice ar fi încetinite dacă oamenii care vorbesc o limbă ar rămâne cu noi, interacționând cu populația generală, amestecându-și vocile cu ale noastre. Mințile simulate ar învăța noi moduri de a vorbi – ne-am schimba cu toții tiparele de limbaj în timp –, dar schimbarea lingvistică ar fi mai lentă dacă vechile generații nu ar muri niciodată. Dacă s-ar realiza transferul minții, am putea vorbi o formă de

[0] H. Palsson și P. Edwards, Seven Viking Romances (Toronto, Canada: Penguin Books, 1985).

proto-indo-europeană și 6.000 de ani mai târziu.

Nu doar limbajul, ci și moda, moravurile, divertismentul, religia și cultura în general se schimbă odată cu generațiile, în ciuda faptului că avem cărți, se pare că luăm mereu lucrurile de la început. Imaginează-ți ce inerție ar avea o cultură dacă oamenii din trecut nu ar muri niciodată. Dinamicile modificărilor culturale ar fi cu totul altele.

Imaginează-ți cum s-ar schimba politica. Filosoful și scriitorul George Santayana a spus: „Cei care nu-și pot aminti trecutul sunt condamnați să-l repete”. 22 Și asta facem. Repetăm trecutul pentru că suferim de amnezie generațională. Memoria politică nu se șterge de tot, dar se estompează în două sau trei generații. Nu vreau să spun că dispare cunoașterea intelectuală. Copiii de școală încă învață fundamentele istoriei, dar caracterul imediat, intensitatea afectivă a memoriei personale, se pierde. În timp, istoria devine teoretică. După lecțiile traumatizante ale celui de-al Doilea Război Mondial, întreaga lume era convinsă de primejdia populismului și a fascismului. Dar oamenii angajați direct în acele evenimente politice au dispărut, urgența evenimentelor de atunci nu s-a transmis bine peste generații și până și cele mai evidente fapte sunt negate de grupuri din ce în ce mai mari de oameni. Populismul și fascismul se insinuează înapoi pe scena politică într-un mod care în urmă cu 50 de ani ar fi fost imposibil.

Sunt sigur că oricine se poate gândi la exemple de greșeli istorice ciclice ale civilizațiilor. Oricât de bine ar conserva cărțile noastre cunoașterea, fiecare nouă generație trebuie să reinventeze arta și să redescopere pentru sine cum să trăiască cu înțelepciune. Ce s-ar întâmpla dacă vechea generație n-ar muri niciodată, ci ar rămâne activă? Ideal, experiența și înțelepciunea politică s-ar acumula și am evita repetarea acelorași greșeli. S-ar putea ca politica să ajungă asemenea științei, în sensul unui progres continuu. Pe de altă parte, pentru fiecare Gandhi pe care ne-am dori să-l avem pe termen nedefinit, am avea de-a face cu un Nero sau un Hitler mult mai greu

de omorât.

Imaginează-ți cum s-ar schimba sistemul universitar. O profesoară își face treaba în mod strălucit timp de treizeci de ani, predând și participând la comisii, și drept urmare primește titlul de profesor emerit. Își continuă lucrul, până când, într-o zi, moare. De la titlul de emerit e promovată la „mortem”. Nu mai are nevoie de un birou fizic, doar de spațiu în platforma virtuală. De acolo predă în continuare și participă la comisii prin videoconferințe, aducându-și contribuția unică. Dacă predă istorie, cunoștințele ei vor deveni deosebit de valoroase pe măsură ce secolele trec, iar noile generații vor vrea să afle despre trecut de la cei care l-au trăit. Să ne închipuim doar cât de multe am ști dacă demnitarii Egiptului antic ar fi printre noi azi și ne-ar povesti despre viețile lor. Să ne imaginăm însă și cât de greu ar fi pentru un profesor tânăr, aflat la început de drum, să obțină o catedră la o universitate ticsită cu generații întregi de savanți morți.

Nu ne gândim de obicei la cei foarte bătrâni ca la niște membri activi ai societății. Sunt prea fragili, prea puțini și prea rămași în urmă. Știu oameni – din ce în ce mai puțini – care nu se descurcă cu un computer și nu știu ce e Internetul. Unii dintre ei nu vor să aibă telefon mobil, pentru că implicațiile sunt prea noi pentru creierul lor. Noi, ceilalți, trăim într-o explozie de invenții tehnologice care pun vechea generație într-un con de umbră. Trăim într-un prezent din ce în ce mai îngust, reprezentat de informația care ne copleșește clipă de clipă. Dar, bun sau rău, transferul minții poate modifica radical acest tipar. Bătrânii nu vor mai fi o specie pe cale de dispariție. Ar fi la fel de prezenți ca toți ceilalți și ar ține la vechiul lor stil de viață. Inovația culturală vine, în general, de la noua generație și este contestată de cea veche. O generație veche care nu ar dispărea ar schimba datele demografice și ar conduce cultura spre o mai mare stabilitate. Viața digitală de apoi s-ar putea să aibă un efect paradoxal asupra societății: pe de-o parte, să o transforme radical, iar, pe de alta, s-o facă mai conservatoare și să-i încetinească ritmul de schimbare.

Nu doar că generațiile anterioare vor rămâne printre noi, ci vor acumula, probabil, și putere. Nu avem niciun motiv să credem că cei vii vor avea vreun avantaj politic, economic sau intelectual față de cei simulați. Gândește-te la slujbele pe care le au oamenii în lumea noastră. Multe necesită acțiune fizică, și tocmai acestea vor fi, probabil, preluate de roboți. Șofer de taxi? Mașinile automate deținute de companii publice sunt aproape o realitate. Măturătorii? Casierii? Muncitorii în construcții? Piloții? Toate slujbele astea vor dispărea probabil pe termen mediu spre lung. Robotica și inteligența artificială le vor înlocui. Pentru celelalte slujbe e nevoie de gândire și, dacă mintea poate fi transferată, pot fi făcute în continuare. Un politician poate lucra din cyberspațiu la fel de bine ca din spațiul real. La fel și un profesor, un manager, un terapeut, un jurnalist, un scriitor sau tipul de la departamentul de reclamații. CEO-ul unei companii, genul Steve Jobs, cu o rețea optimă de conexiuni neurale, care îl fac excepțional de bun la ce face, poate să fie manager fără să fie prezent fizic. Dacă trebuie să dea mâna cu cineva, poate să împrumute un robot umanoid de la un serviciu de închiriat roboți și să petreacă astfel câteva ore în lumea reală, întâlnindu-se cu oameni și socializând. Până și denumirea de lume „reală” mi se pare discriminatorie. Ambele lumi ar fi la fel de reale. Poate că ar fi mai potrivite denumirile lumea „fundație” („*foundation*” *world*) și lumea „nor” („*cloud*” *world*)

Lumea „fundație” ar fi plină de oameni foarte tineri – sub 80 de ani –, care încă ar acumula experiențe valoroase. Datoria lor nerostită ar fi să se dezvolte, să se pună la curent cu ce e nou și să acumuleze înțelepciune și experiență înainte de a se alătura celor din lumea „nor”. Centrul puterii și al culturii s-ar muta rapid în lumea „nor”. Nici nu ar putea să fie altfel. Acolo se vor aduna cunoașterea, experiența și legăturile politice. În acest scenariu, lumea „fundație” ar fi un fel de lume larvară a minților imature, iar lumea „nor” ar fi locul în care viața ar începe cu adevărat.

Să vorbim acum despre călătoria în spațiu.

Star Trek ne-a oferit prima viziune optimistă, populară, a viitorului: omenirea va eradica războiul și boala și se va răspândi în galaxie, ca specie pașnică și curioasă din punct de vedere științific, cu ajutorul navelor spațiale cu motoare cu distorsiune. Acolo vom anihila cu discreție speciile de extratereștri care ne incomodează. Fie că împărtășești această viziune a viitorului sau o alta din meniul de opțiuni științifico-fantastice, există în cultura noastră predispoziția spre a crede că viitorul nostru se leagă de spațiu și că specia noastră va coloniza galaxia.

Dar călătoria în spațiu e mult mai complicată decât se crede. Corpul uman nu este compatibil cu alcătuirea toxică a spațiului. Chiar dacă sunt protejați de o cabină cu aer oxigenat, oamenii tot se vor confrunta cu o ploaie letală de raze cosmice care va trece prin aproape orice fel de scut. Stația Spațială Internațională care e în prezent pe orbită e protejată de câmpul magnetic al Pământului, care refractează razele cosmice. Dacă se vor duce pe Marte, mult dincolo de influența protectoare a Pământului, oamenii s-ar putea să moară. Să ajungă dincolo de Marte e o provocare exponențial mai dificilă.

Chiar dacă am rezolva problemele tehnice care țin de capacitatea oamenilor de a trăi în spațiu, tot rămâne problema timpului. Cel mai apropiat sistem solar, Alpha Centauri, e la o distanță mai mare de patru ani-lumină. Călătoria cu 10% din viteza lumii, mult peste capacitatea oricărei tehnologii spațiale din prezent, ar dura numai la dus 45 de ani. Dacă am inventa o navă spațială care printr-un miracol ar călători cu 99,9% din viteza lumii, tot ne-ar lua mai mult de 100.000 de ani să traversăm galaxia. Durata vieții umane e prea scurtă pentru o explorare semnificativă a spațiului. Vom avea, poate, o călătorie spațială transgenerațională. Dar, în cazul ăsta, provocările tehnice legate de prelungirea vieții biologice nu ar face decât să se complice.

Visul speciei noastre de a se răspândi glorios în galaxie nu se va realiza niciodată. Un asemenea viitor rămâne

iluzoriu. Literatura științifico-fantastică greșește în privința asta.

Există însă o soluție pentru explorarea spațiului. Mințile transferate nu au nevoie de oxigen, de presiune atmosferică sau de hrană organică. Nu au nevoie de un mediu asemănător celui de pe Pământ încapsulat într-o cabină. De asemenea, pot să trăiască pe termen nedefinit și pot cu ușurință să-și accelereze sau să-și încetinească ritmul experienței conștiente. Să ne imaginăm un grup de oameni – vorbim de sute, de mii sau chiar de copii ale tuturor minților care vor fi fost vreodată transferate – aflat într-o platformă simulată dintr-o navă spațială. Sarcina lor e să exploreze. Pentru că trăiesc într-o lume virtuală, nu stau înghesuiți. Se pot răspândi confortabil într-un oraș întreg, sau într-o copie virtuală a Pământului, sau într-un alt peisaj simulat. Dacă vor, pot să creeze o navă *Enterprise* și să trăiască acolo exact ca în serial, cu ecrane mari pe punte, care să le arate ce progrese fac, și cu butoane care licăresc ca să-i amuze. Oamenii de știință și tehnicienii ar monitoriza universul din jur prin ecranul calculatorului și cu ajutorul instrumentelor, alegând în consecință traiectoria zborului. Dacă ar vrea să evite o perioadă plictisitoare de câteva secole între două sisteme solare, ar putea pur și simplu să rotească un ac pe un cadran ca să-și încetinească viteza de procesare, făcând să pară că a trecut doar o jumătate de oră – suficient timp pentru a ajusta traiectoria atât cât e nevoie.

După ce ajung într-un sistem solar extraterestru, dacă vor să aterizeze pe o planetă pe care să o exploreze, nu au nevoie de atmosferă respirabilă. Tot ce le trebuie e o temperatură care să nu le topească echipamentul și o gravitație care să nu-i zdrobească. Dacă ai face parte din echipa aceea, în loc să-ți pui un costum de astronaut și să ieși dintr-o capsulă, ai putea să-ți transferi rețeaua neurală într-un navigator robotic și să te-aventurezi în noul peisaj, colectând probe și făcând poze. La sfârșit, te-ai întoarce la Pământul simulat sau la orice mediu ar oferi nava-mamă.

Presupun că majoritatea oamenilor de azi ar prefera ca

universul să fie colonizat de oameni reali. O adevărată civilizație spațială nu se poate limita la sonde virtuale și robotice. Vrem pe cineva ca Neil Armstrong să iasă din navă și să pună piciorul pe o nouă planetă. Dar cred că singura noastră componentă esențială, partea care ne definește, este mintea – și într-o zi tehnologia îi va permite minții umane să călătorească liberă prin univers.

Cred că vorbim de un proces inevitabil, în trei pași. Mai întâi va fi inventat transferul minții. Motivația psihologică, dorința de a conserva mintea după moartea biologică, este atât de mare, încât oamenii de știință și inginerii se îndreaptă deja spre acest scop. Următorul pas logic este să trimitem mințile umane transferate pe platforme artificiale în locuri în care oamenii nu pot ajunge fizic. Drept urmare, vom crea o civilizație care călătorește pe mari distanțe în spațiu, pe durate foarte lungi, și care se va răspândi în toată galaxia. Cheia acestui viitor este înțelegerea faptului că mintea umană este informație și că, în principiu, poate fi transferată dintr-un creier fizic într-un sistem artificial. Momentul în care vom deveni o civilizație care călătorește în spațiu va fi momentul în care vom înțelege ce este conștiința dintr-o perspectivă tehnică.

Vreau să mai spun ceva despre transferul minții, ceva legat de socialitatea umană. Oamenii sunt atrași unii de alții.

Conectivitatea face parte din natura noastră. Da, sunt destui singuratici, sihaștri sau agorafobi, dar cei mai mulți dintre noi comunică cu o rețea de oameni. Nu ne-a fost de ajuns să ne scriem scrisori, așa că am inventat telefonul, apoi e mailul, apoi SMS-urile. Acum suntem nedezlipiți de telefonul inteligent și trăim prin media sociale. Dacă am avea opțiunea telepatiei, mulți ar adopta-o rapid și ar alege să fie în contact continuu cu o rețea de miliarde de alți oameni.

Orice ar zice adepții magiei sau ai pseudoștiinței, oamenii nu și-au împărtășit niciodată direct gândurile, în principal pentru că nu ne naștem cu porturi USB în cap. Dar

o minte simulată ar fi alcătuită din informații care sunt gestionate și manipulate într-o platformă artificială. Din punct de vedere tehnic, o minte simulată ar putea comunica direct cu o alta.

Nu știu cum ar funcționa o astfel de tehnologie. Frumusețea și simplitatea transferului minții este că tehnicienii nu trebuie să înțeleagă în detaliu cum funcționează creierul; tot ce trebuie să facă e să-l copieze. Dacă e realizată o copie suficient de fidelă, ar trebui să funcționeze la fel ca originalul. Dar conectarea directă a unei minți la o alta ar implica o înțelegere mai profundă a modului în care sunt procesate gândurile. Ce outputuri folosești? Ce imputuri? O abordare rudimentară nu ar face decât să preia la întâmplare semnale emise de mintea unei persoane și să le transmită ca zgomot fără sens unei a doua persoane. Date fiind limitările cunoașterii neuroștiințifice din prezent, nici nu pot să-mi imaginez cum am putea construi o astfel de tehnologie. Îmbinarea a doua minți transferate este rezervată viitorului îndepărtat, când creierul uman va fi mult mai bine înțeles. Dacă totuși această fuziune se va realiza vreodată, vom avea de-a face cu o schimbare fundamentală a dinamicii fluxului informatic. În loc să fie o colecție de indivizi simulați, mințile transferate vor deveni un nexus de inteligență în care individualitatea va dispărea. Oamenii care încă vor trăi în lumea biologică vor fi slabi din punct de vedere mintal, biete fragmente dispensabile de creier, preocupate să acumuleze experiență, sperând la șansa de a muri și la privilegiul de a se topi în inteligența colectivă.

Astfel de versiuni ale viitorului ne sperie de obicei – și pe mine mă sperie destul de tare –, dar cred că repulsia e provocată în principal de frica de necunoscut. Transferul minții deschide atât de multe posibilități pentru viitor, încât nu știm cum să reacționăm.

Îmi place să-mi închipui un grup de oameni de Cro-Magnon, bărbați și femei, care stăteau noaptea în jurul focului în urmă cu vreo 30.000 de ani. Unul dintre ei spune o poveste despre un viitor ciudat. Lumea lor va fi nivelată,

pădurile tăiate, sălbăticia acoperită de pietre artificiale în formă de cuburi suprapuse care se ridică până la cer. Magnificele animale pe care le vânează și pe care le pictează pe pereții peșterilor vor fi toate dispărute – bourii, mamuții, urșii de peșteră. Viața spirituală a vânătorului va fi dispărut. Nimeni nu va mai ști să facă cu mâinile lui sulițe, sau săgeți, sau orice altceva. Cei mai mulți oameni stau înăuntru, la o lumină slabă, apăsând toată ziua mici butoane care licăresc pe plăci dreptunghiulare; trupurile le sunt moi și lente. Aerul e sufocat de vapori artificiali, afară lumea vuiește într-un vacarm asurzitor și neîntrerupt. Îmi închipui că ceilalți membri ai clanului ar strâmba din nas și-ar condamna această imagine distopică.

Și totuși aici suntem și nu ne deranjează. Puțini dintre noi ar vrea să se întoarcă la un stil de viață precivilizat, de vânători-culegători. Știm că lumea noastră are problemele ei, dar nu vrem să o schimbăm prea mult. Când ne imaginăm un viitor utopic, ne gândim la o lume asemănătoare celei de azi, din care excludem însă problemele cele mai grave, adăugând o serie de progrese tehnologice. Orice viziune a unui viitor prea diferit de prezent trebuie, prin definiție, să fie distopică. Cred că, indiferent de perioada în care au trăit, oamenii au avut aceleași sentimente legate de viitor.

Dar, fie că ne place sau nu, lumea se va schimba. Inteligența artificială și transferul minții o vor restructura. Oamenii din viitor se vor descurca probabil la fel cum ne descurcăm și noi în lumea noastră. Cel mai probabil, vor fi prea obișnuiți cu viața lor ca să vrea să se întoarcă la felul nostru de a trăi, haotic și plin de limitări.

Când mă gândesc cât mai departe în viitor, schimbarea esențială pe care o văd – momentul-cheie din istoria speciei noastre – este clipa în care oamenii vor înțelege conștiința. Dacă vom reuși să înțelegem conștiința dintr-o perspectivă pragmatică, tehnică, vom putea avea un viitor remarcabil. În acel viitor, mintea e ceva prețios, ce trebuie îngrijit, dezvoltat și apoi salvat, ceva ce poate fi prelevat din platforma biologică inițială și transferat, multiplicat,

ramificat, menținut pe termen nedefinit, poate chiar combinat cu alte minți. Văd cum mintea e decuplată de la biologie și văd cum dispare granița dintre inteligența artificială și inteligența umană. Și mai văd cum această proprietate transcendentă a minții se răspândește în spațiu și explorează galaxia, timp de milioane de ani. Tehnologia minții este poate cea mai bună șansă pe care o avem spre un viitor promițător.

ANEXĂ
Cum se construiește conștiința vizuală

Să facem un experiment mintal. Scopul nostru e să înțelegem în detaliu cum să construim o mașinărie care vede un măr și e conștientă de ceea ce vede. Acest exercițiu are două aplicații potențiale. Prima este aceea de a explica teoria schemei atenției. Logica aflată la baza teoriei va fi expusă în cea mai simplă formă. A doua constă în a le arăta tehnicienilor calea spre conștiința artificială.

Folosesc adesea un măr în exemplele mele pentru că întrunește două aspecte vizuale comune: forma simplă și culoarea intensă. Vom înzestra mașinăria cu o cameră-ochi și cu un computer-creier și vrem să aflăm ce trebuie să construim în creierul artificial pentru a-l face să susțină că are o experiență subiectivă, conștientă a mărului. Acest experiment de gândire este supus unei constrângeri: tot ce punem în mașinăria noastră ipotetică trebuie să poată fi construit. Unele componente s-ar putea să existe deja. Altele nu există încă, dar sunt plauzibile, luând în considerare unde a ajuns tehnologia astăzi, așa că avem voie să le includem.

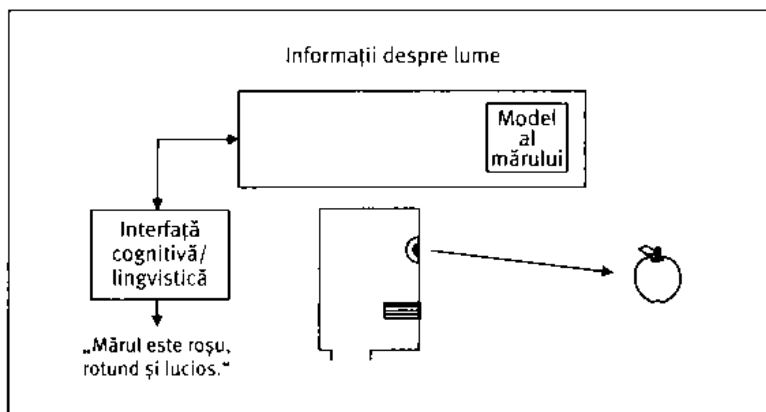
Voi prezenta aici trei versiuni diferite ale mașinăriei noastre, fiecare puțin mai avansată decât cea anterioară, și în cazul versiunii finale voi susține că mașinăria are în esență același tip de conștiință pe care îl avem noi, oamenii. Nu vreau să spun că mașinăria are o viață

interioară bogată. Nu ar avea conștiință de sine, emoții, imaginație sau planuri. Ar fi conștientă de un singur element – de măr. Și totuși această mică felie de conștiință ne va arăta calea spre a construi o conștiință cu un conținut mai bogat.

Figura A.l prezintă un robot care se uită la un măr. Căsuța de deasupra capului robotului reprezintă informația prezentă în dispozitiv, iar în această primă diagramă conține un singur element: informație vizuală despre măr.

Procesarea vizuală este unul dintre subiectele cele mai studiate în neuroștiințe. Până acum a fost identificat în linii mari modul în care funcționează sistemul vizual uman, chiar dacă unele enigme și detalii sunt încă neelucidate. Tot ce trebuie să știm pentru proiectul nostru de construire a unui creier este că au fost deja realizate sisteme vizuale artificiale, chiar dacă simple și limitate în comparație cu creierul uman.[0] Îi vom oferi robotului nostru capacitatea de a asimila o imagine vizuală printr-o cameră și de a construi o serie bogată și detaliată de informații despre măr. Sistemul vizual artificial aduce laolaltă informații privind culoarea, forma, dimensiunea și poziția mărului, construind ceva asemănător unui dosar, o serie de informații care se schimbă constant pe măsură ce sunt preluate noi semnale. Acest pachet de informații e numit uneori model intern. E un fel de simulare a unui măr.

[0] R. Klette, Concise Computer Vision (New York: Springer, 2014); L.G. Shapiro și G.C. Stockman, Computer Vision (Upper Saddle River, NJ: Prentice Hall, 2001); M. Sonka, V. Hlavac și R. Boyle, Image Processing, Analysis, and Machine Vision (Stamford, CT: Cengage Learning, 2008).



A.1. Un robot recepționează informație vizuală printr-o cameră și construiește un model intern sau un pachet de informații despre măr. Un motor de căutare („interfața cognitivăzlingvistică”) îi permite mașinăriei să răspundă la întrebări pe baza informației interne.

În creierul real, biologic, aceste simulări nu sunt foarte precise. Creierul ar irosi energie și resurse de procesare realizând o descriere a mărului detaliată, precisă din punct de vedere științific. Culoarea mărului, de exemplu, e în parte un construct al creierului. În realitate, mărul are un spectru de refracție, nu o culoare. Ochii și creierul simplifică acest spectru și-i atribuie mărului o culoare. Culoarea este o caricatură, o aproximație făcută în grabă. Pentru supraviețuire, eficiența este totul. Nu vrei să procesezi atât de profund încât să nu poți reacționa îndeajuns de repede.

Până aici, nimic din ce-am spus nu contrazice premisele cu care am pornit la drum. Dacă am avea o cameră și un computer, am putea construi un sistem precum cel din Figura A.1. Dar este robotul nostru *conștient de* măr?

Potrivit unei definiții a cuvântului, da. Cercetătorii conștiinței folosesc uneori termenul de *conștiință obiectivă* pentru a desemna informația asimilată și procesată.[0] Da,

---

[0] P.M. Merikle, D. Smilek și J.D. Eastwood, „Perception without awareness: Perspectives from Cognitive Psychology”, Cognition, 79 (2001): 115 – 134; R. Szezepanowski și L. Pessoa, „Fear Perception: Can Obiective and Subiective Awareness Measures Be Dissociated?”, Journal

mașinăria din Figura A.1 este conștientă obiectiv de măr. Conține informația.

Dar este conștientă *subiectiv*? Are o experiență subiectivă a culorii roșii, a rotunjimii și a strălucirii, așa cum am eu sau cum ai avea tu dacă ne-am uita la un măr? Unii oameni de știință încă susțin că da, conștiința e ceea ce simți când procesezi informație.[0] Întrucât mașinăria procesează mărul, este în mod necesar conștientă subiectiv de măr. Eu numesc acest tip de teorie ideea „informației adiacente”: conștiința este un produs secundar inevitabil al procesării informației, așa cum căldura este un produs secundar al circuitelor electrice. Din această perspectivă, dacă construiești o mașină de calcul, conștiința apare oricum. Dacă e adevărat, atunci am terminat. Am reușit să creăm diagrama unei mașinării conștiente, deși conștiința rămâne în continuare neexplicată.

Dar nu cred că mașinăria e gata. Să întrebăm robotul. Dacă tot facem acest experiment de gândire, putem și să adăugăm o interfață lingvistică, un motor de căutare asemenea lui Siri (numit în Figura A.1 „interfață cognitivă/lingvistică”). Primește întrebări, caută în baza de date disponibilă și răspunde. Dat fiind că poate fi construită cu tehnologia de azi, avem voie să includem acest motor de căutare în robotul nostru, chiar dacă această utilizare a lui este complet nouă.

Întrebăm mașinăria:

— Ce-i aici?

Mașinăria răspunde:

— Un măr.

— Care sunt proprietățile mărului?

— E roșu, rotund, lucios, turtit în partea de sus, se află acolo…

Poate răspunde la aceste întrebări de bază pentru că

of Vision 10 (2007): 1-17.

[0] M. Tegmark, „Consciousness as a State of Matter”, arxiv (2014): 1401.1219; G. Tononi, M. Boly, M. Massimini și C. Koch, „Integrated Information Theory: From Consciousness to its Physical Substrate”, Nature Reviews Neuroscience 17 (2016): 450 – 461.

deține informațiile relevante. Mai mult, răspunsurile ei sunt bogate și flexibile, fiindcă are o cunoaștere a mărului bazată pe un model, iar modelul pe care îl consultă este o descriere bogată, chiar dacă nu întru totul precisă, a mărului. Avem acum o mașinărie care poate să vadă un măr, să-l proceseze și să facă afirmații explicite despre el. Construcția ei este perfect posibilă din punct de vedere tehnic cu mijloacele de azi.

Și, cu toate astea, ca teorie a conștiinței, Figura A.1 este incompletă. Ca să demonstrăm asta, hai să-i punem mașinăriei următoarea întrebare evidentă: „Ești tu conștientă de măr?"

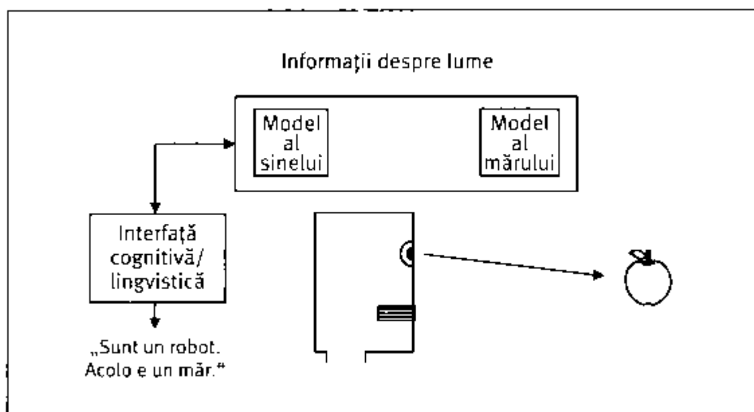
Motorul de căutare investighează modelul intern și nu găsește niciun răspuns la această întrebare. Găsește o mulțime de informații despre măr, dar nimic despre conștiință – despre ce e conștiința sau dacă ea are această calitate ciudată. Îi lipsesc și informații despre sine. La urma urmei, am întrebat: „Ești *tu* conștientă de măr?" îi lipsesc informații despre entitatea numită „tu". Întrebarea noastră nu are sens pentru ea. În cel mai bun caz poate răspunde: „Nu pot calcula". Putem la fel de bine să întrebăm o cameră digitală: „Ești tu conștientă de fotografia pe care tocmai ai făcut-o?" Nimic din ce-am pus în mașinărie până acum nu o poate face să afirme că are conștiință.

Ține de logica fundamentală faptul că o mașinărie care procesează informație nu poate face o afirmație – nu are output informațional – decât dacă deține informațiile necesare afirmației. Mașinăria din Figura A.l nu conține nicio informație despre ce este conștiința și cu atât mai puțin despre propria conștiință. Nu putem spera că, fiindcă are oarece complexitate, mașinăria va susține dintr-odată că e conștientă. Așa cum am construit-o, mașinăria nu conține nicio informație despre zebre, de exemplu, așa că nu are cum să înceapă brusc să vorbească despre zebre. Tehnic vorbind, speranța că mașinăria va spune pe neașteptate: „Sunt conștientă" este o încălcare a logicii. Va trebui să includem mai multe informații ca să extindem gama de răspunsuri pe care le poate oferi.

Figura A.2 prezintă a doua versiune a proiectului de construire a unui creier. Am adăugat aici un al doilea model intern: un model al sinelui. Din nou, modelul este compus dintr-o serie de informații reunite în creier. Informația poate descrie forma fizică și structura corpului, așa-numita schemă corporală.[0] Poate include memorie autobiografică, aceasta jucând un rol important în orice model psihologic al sinelui.[0] Știm că creierul uman conține un model al sinelui, probabil foarte complex și probabil extins în multe arii cerebrale. Îi putem atribui robotului nostru un model al sinelui, dat fiind că e fezabil din punct de vedere tehnic să-i oferim o serie arbitrară de informații.

[0] M.S.A. Graziano și M. Botvinick, „How the Brain Reprezents the Body: Insights From Neurophysiology and Psychology", în W. Prinz și B. Hommel (coord.), Common Mechanisms în Perception and Action: Attention and Performance XIX (Oxford, UK: Oxford University Press, 2002), 136 – 157; N. Holmes și C. Spence, „The Body Schema and the Multisensory Representation (s) of Personal Space", Cognitive Processing 5 (2004): 94 – 105; F. De Vignemont, Mind the Body: An Exploration of Bodily Self-Awareness (Oxford, UK: Oxford University Press, 2018).

[0] S. Bluck și T. Habermas, „The Life Story Schema", Motivation and Emotion 24 (2000): 121 – 147; M.A. Conway și C.W. Pleydell-Pearce, „The Construction of Autobiographical Memories în the Self-Memory Sistem", Psychological Review 107 (2000): 261 – 288; M.A. Conway, J.A. Singer și A. Tagini, „The Self and Autobiographical Memory: Correspondence and Coherence", Social Cognition 22 (2004): 491 – 529.



A.2 Robotul a primit un al doilea model intern, un pachet de informații care îl descriu pe el.

Acum îi spunem mașinăriei îmbunătățite:

— Cine ești? Vorbește-ne despre tine.

Datorită noului model intern, mașinăria are informațiile necesare pentru a răspunde. Spune:

— Sunt un om metalic; am 1,80 metri înălțime; pot să-mi îndoi mâinile de la cot și umăr; m-am născut într-un laborator; a fost ziua mea acum cinci zile…

Motorul de căutare poate accesa și împărtăși informațiile disponibile în modelul intern al sinelui. Parte din aceste informații s-ar putea să fie despre mașinăria însăși, iar altele, despre evenimente din trecut care o definesc.

Este robotul conștient acum? Unii ar spune că da. Conform multor teorii, conștiința e echivalentă cunoașterii de sine. Unii oameni de știință care susțin această perspectivă se concentrează asupra schemei corporale. Potrivit acestei abordări, cel mai bazal, cel mai primitiv tip de conștiință este cunoașterea sinelui fizic și a modului în care se mișcă.[0] Știu că părțile corpului meu sunt ale mele:

---

[0] O. Blanke, „Multisensory Brain Mechanisms of Bodily Self-Consciousness”, Nature Reviews Neuroscience 13 (2012): 556 – 571; O. Blanke și T. Metzinger, „Full-Body Illusions and Minimal Phenomenal Selfhood”, Trends în Cognitive Sciences 13 (2009): 7 – 13; C. Preston, B.J. Kuper-Smith și H.H. Ehrsson, „Owning the Body în the Mirror: The Efect of Visual Perspective and Mirror View on the Full-Body Illusion”,

fac parte din mine și sunt fundamental separate de celelalte obiecte din jurul meu. Știu unde mă aflu în spațiu: sunt aici. Îmi cunosc perspectiva: aud, și văd, și înțeleg lumea din punctul în care mă aflu fizic. Alți oameni de știință se concentrează asupra cunoașterii psihologice mai abstracte a sinelui.[0] Conform acestei perspective, sunt conștient pentru că îmi cunosc traiectoria în viață, îmi știu motivațiile și planurile. Pot să construiesc o narațiune fluidă pentru a spune cine sunt și de ce fac lucrurile pe care le fac.

Mașinăria din Figura A.2 deține multe informații despre sine. Tehnica ne permite să construim orice volum de cunoaștere de sine – un element desigur important pentru conștiință. Dar, iarăși, cred că diagrama oferă o imagine incompletă. Ca să demonstrăm asta, hai să întrebăm mașinăria:

— Care e relația mintală dintre tine și măr?

Mașinăria se blochează. Motorul de căutare se raportează la modelul intern și găsește multe informații despre sine, multe informații despre măr, dar nimic despre relația mintală dintre cele două – nicio informație despre ce înseamnă o relație mintală. Nici nu poate să proceseze întrebarea. Echipat doar cu elementele oferite până acum, robotul nu poate răspunde la întrebare. Nu poate afirma că e conștient de măr, și conștiința e irelevantă pentru el.

Avantajul modelelor interne este că monitorizează și fac predicții despre lucrurile importante din lume. Universul robotului nostru conține două elemente evidente: mărul și robotul însuși. I-am oferit așadar două modele interne: un model al mărului și un model al sinelui. Dar am trecut cu vederea un al treilea element, subtil, din lumea lui: relația operantă dintre robot și măr. Pentru a-i oferi robotului o

---

Scientific Reports 5 (2015): 18345.

[0] M.S. Gazzaniga, The Bisected Brain (New York: Appleton Century Crofts, 1970); R.E. Nisbett și T.D. Wilson, „Telling More Than We Can Know-Verbal Reports on Mental Processes”, Psychological Review 84 (1977): 231 – 259.

descriere completă a lumii sale trebuie să-l echipăm cu un al treilea model intern, care să-i descrie procesul de atenție vizuală.

În cazul unui om, atenția vizuală este necesară pentru că un decor tipic poate include un măr, o farfurie, o masă, un scaun – uneori sute de lucruri, prea multe pentru a putea fi procesate în profunzime. Creierul trebuie să prioritizeze, să-și concentreze resursele asupra unei serii limitate de obiecte la un moment dat. Doar unele elemente câștigă competiția într-un moment anume, modele interne ale acestora transmițând semnale suficient de puternice pentru a avea un efect asupra sistemelor cerebrale.[0] Elementele trecute cu vederea rămân neprocesate – nici măcar nu le înregistrăm prezența –, în timp ce obiectele care primesc atenție sunt procesate în profunzime. Creierul poate extrage detalii și sensuri din elementele selectate și poate decide cum să reacționeze.

Uneori, când oamenii folosesc colocvial cuvântul *atenție*, se referă doar la obiectul aflat în centrul de interes. De exemplu, o conversație poate fi centrul meu de interes, atrăgând cea mai mare parte a resurselor mele mintale, în timp ce mărul pe care îl mănânc are o importanță marginală. Dar, în sens științific, atenția înseamnă ceva mai mult. În acest exemplu, și conversația, și mărul primesc atenție. Reprezentările lor din creier sunt amplificate îndeajuns pentru a fi procesate într-un mod semnificativ. Nici măcar nu știm că sunt prezente sutele de obiecte, sau gânduri, sau semnale senzoriale neluate în seamă. Când vorbesc despre atenție, mă refer la relația mai cuprinzătoare dintre creier și obiectele asupra cărora acesta își concentrează resursele.

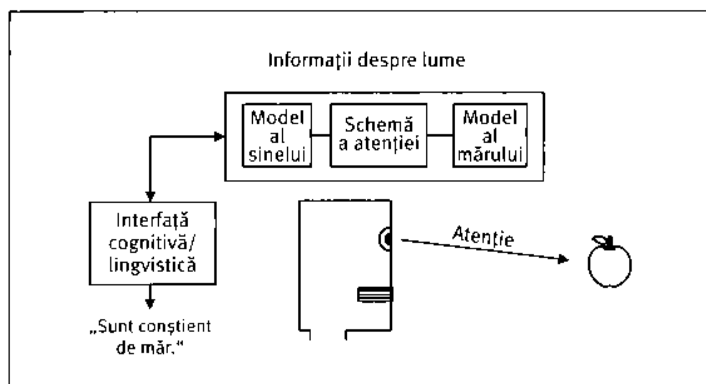
Atenția vizuală a fost construită artificial, deși sistemele artificiale sunt mult mai simple decât versiunea biologică,

---

[0] D.M. Beck și S. Kastner, „Top-Down and Bottom-Up Mechanisms în Biasing Competition în the Human Brain”, Vision Research 49 (2009): 1154 – 1165; R. Desimone și J. Duncan, „Neural Mechanisms of Selective Visual Attention”, Annual Review of Neuroscience 18 (1995): 193 – 222.

umană.[0]9 Dat fiind că există deja mașinării dotate cu atenție vizuală, avem voie să includem această funcție în proiectul nostru de construire a unui creier.

Figura A.3 prezintă versiunea a treia și ultimă a robotului. Aici, robotul își îndreaptă atenția vizuală spre măr. Să observăm că este dotat cu trei modele interne, care îi oferă o descriere completă a celor trei componente principale ale lumii sale.



A.3 Principalele componente ale teoriei schemei atenției. Robotul conține un model al mărului, un model al sinelui și un model al relației atenționale dintre sine și măr.

Un model descrie mărul, unul descrie sinele și unul descrie această a treia componentă invizibilă, relația atențională dintre sine și măr. Acest model al atenției se numește „schemă a atenției”, analogă schemei corporale, care descrie corpul.

Știm deja ce informații trebuie să conțină modelul mărului și modelul sinelui. Dar ce informații trebuie să conțină modelul atenției? Cum am descrie atenția în mod

[0] G. Deco și E.T. Rolls, „Neurodynamics of Biased Competition and Cooperation for Attention: A Model with Spiking Neurons”, Journal of Neurophysiology 94 (2005): 295 – 313; L. Layon și S.L. Denham, „A Biased Competition Computational Model of Spațial and Obiect-Based Attention Mediating Active Visual Search”, Neurocomputing 58 (2004): 655 – 662; J. Reynolds și D. Heeger, „The Normalization Model of Attention”, Neuron 61 (2009): 168 – 185; J.K. Tsotsos, A Computational Perspective on Visual Attention (Cambridge, MA: MIT Press, 2011).

util?

Descrierea poate fi atât de simplă: „Mărul este ceea ce mă preocupă”; altfel ar fi un model gol, o linie punctată între celelalte două modele. O schemă a atenției are nevoie de informații despre atenția în sine – despre ce înseamnă concentrarea atenției. Dar, la fel ca modelul intern simplificat al mărului, o schemă bună a atenției nu ar risipi resurse pe informații despre detaliile tehnice microscopice. În cazul creierului real, printre detaliile tehnice ale atenției se numără neuronii, sinapsele și semnalele electrochimice aflate în competiție. În cazul robotului nostru, detaliile pot include cabluri, cipuri de procesare și porți logice de siliciu. Sistemul nu trebuie să cunoască toate amănuntele. Mașinăria noastră are nevoie de o descriere de suprafață, săracă în detalii, a atenției pe care-o acordă mărului. Ar putea descrie atenția ca pe un mod de a poseda ceva cu mintea, ca pe-o modalitate de-a capta informația și de-a o cunoaște în profunzime. Ar putea descrie atenția și ca pe un proces în mare parte intern. Ar putea de asemenea să se refere la consecințele sale previzibile: atenția conferă capacitate de reacție, amintiri, puterea de a lua decizii.

Să-i punem acum o întrebare mașinăriei din Figura A.3:

— Care e relația mintală dintre tine și măr?

Motorul de căutare accesează modelele interne ale mașinăriei și răspunde pe baza informațiilor disponibile. Spune:

— Mărul se află în posesia mea mintală.

Promițător răspuns, dar putem să sondăm mai în profunzime:

— Spune-ne despre această posesie mintală. Care sunt atributele ei fizice?

Pentru a ne asigura că mașinăria înțelege întrebarea, adăugăm:

— Știi ce sunt acelea atribute fizice?

Mașinăria deține o schemă corporală, care îi descrie corpul fizic, precum și un model intern al mărului, care descrie un obiect fizic. Așa că poate să spună:

— Da, știu ce sunt proprietățile fizice. Soliditatea,

greutatea, impulsul, opacitatea – acestea sunt proprietățile obiectelor.

Dar, pe baza informațiilor pe care le are la dispoziție, poate spune (dacă are un vocabular bogat):

— Posesia mintală asupra mărului nu are *în sine* proprietăți fizice care pot fi descrise. Deși există, nu are soliditate, greutate, impuls sau opacitate. Nu există în acești parametri. Are însă o esență non-fizică. Într-un sens, e *metafizică*. Dar poate fi localizată fizic – există undeva înăuntrul meu. Face parte din mine – este acea parte metafizică din mine care îmi permite să înțeleg cu mintea. Nu procesez doar faptul că mărul e roșu; intru în *posesia* mintală a culorii roșii.

Adăugând o schemă a atenției, am construit o mașinărie care afirmă că are o conștiință suspect de asemănătoare cu conștiința pe care oamenii susțin că o au.

Fiindcă am construit robotul chiar noi și pentru că știm cum funcționează, rămânem sceptici. Continuăm cu întrebările:

— Mulțumesc pentru explicație, dar nu ești altceva decât o mașinărie și *era de așteptat* să emiți toate informațiile astea. Nu faci decât să accesezi un model intern, schema atenției cu care ești înzestrat, și să raportezi informația pe care o găsești acolo. E, desigur, impresionant că poți da răspunsuri atât de flexibile și de bogate accesând cunoașterea bazată pe modelele tale interne. Cu toate astea, nu faci decât să raportezi informațiile pe care le ai.

Auzindu-ne obiecția, motorul de căutare accesează modelele interne, dar nu găsește nimic care să se potrivească descrierii. Să nu uităm că modelele sale interne sunt prea sărace în detalii pentru a descrie modul în care mașinăria funcționează de fapt. Date fiind informațiile limitate care îi sunt disponibile, mașinăria răspunde:

— Nu știu nimic despre modele interne, informații sau operații. Iată ce știu: mărul e acolo, eu sunt aici și am conștiința mărului. Conștiința în sine nu e ceva fizic. Nu e o operație sau un model, orice ar fi acelea. E o calitate non-

fizică pe care o dețin și care face ca mărul să fie prezent pentru mine și care îmi permite să reacționez.

Mașinăria nu doar că susține că e conștientă, ci, la fel ca majoritatea oamenilor, respinge explicația mecanicistă a propriei existențe și crede într-o teorie metafizică. Așa e construită, condiționată de informațiile interne cu care a fost înzestrată. Atâta timp cât se bazează pe introspecție – și prin asta mă refer la motorul cognitiv de căutare care accesează modelele interne ale mașinăriei –, va apela întotdeauna la aceeași descriere metafizică a sinelui.

Se cunosc destul de multe despre vederea artificială pentru ca acest sistem limitat să poată fi construit cu tehnologia de azi, cel puțin într-o formă simplă. Va fi greu să abordăm și alte domenii până la noi progrese tehnologice, dar, în principiu, aceeași logică se poate aplica oricărui element. Am putea, teoretic, să înlocuim mărul cu un sunet, cu o atingere, cu o amintire, cu o emoție sau cu gândul că *2* plus *2* fac 4. Am putea construi diagrama gândului „Sunt conștient de mine însumi” sau chiar „Sunt conștient că sunt conștient”. Conștiința privind mărul este doar începutul.